
Inferring the structure of dwarf ellipticals with a novel model selection framework

Mathias Lipka



München 2024

Inferring the structure of dwarf ellipticals with a novel model selection framework

Mathias Lipka

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Mathias Lipka
aus Landshut

München, den 11.09.2024

Erstgutachter: P.D. Dr. Roberto Saglia
Zweitgutachter: Prof. Dr. Andreas Burkert
Tag der mündlichen Prüfung: 28.11.2024

Contents

List of Figures	v
Zusammenfassung	vii
Abstract	ix
1 Introduction	1
1.1 The ETG-LTG dichotomy	2
1.2 Dwarf ellipticals	4
1.2.1 Dwarf ellipticals - A distinct subclass in the ETG sequence?	4
1.2.2 Dwarf ellipticals - Remnants of spiral galaxies?	8
1.3 The observables	14
1.3.1 Photometry	14
1.3.2 Spectroscopy	16
1.4 Dynamical Modelling	20
1.4.1 Jeans equations and the mass-anisotropy degeneracy	21
1.4.2 Schwarzschild models	24
1.5 The Basics of Model Selection and Information Theory	29
1.6 Goals and outline of this thesis	32
2 A novel approach to optimize the regularization and evaluation of dynamical models using a model selection framework	35
3 The VIRUS-dE Survey I: Stars in dwarf elliptical galaxies - 3D dynamics and radially resolved stellar initial mass functions	65
4 The VIRUS-dE Survey II: Cuspy and round halos in dwarf ellipticals - A result of early assembly?	123
5 Conclusions	171
A A simple data-driven method to optimize the penalty strengths of penalized models and its application to non-parametric smoothing	175

Acknowledgements

195

List of Figures

1.1	The Hubble tuning fork	2
1.2	Mean surface brightness correlations of ETG galaxies	6
1.3	Sersic index vs absolute magnitude	7
1.4	The Luminosity functions of E and dE galaxies	8
1.5	Mean surface brightness correlations together with late-type galaxies	10
1.6	Faint spiral arm structure in the R-band image of a dE	11
1.7	HST image of VCC 2048	12
1.8	Illustration of discy and boxy isophotes	13
1.9	Morphological parallel-sequence of galaxies	13
1.10	dEs in Virgo	15
1.11	Example of a kinematic fit to a spectrum	17
1.12	Gauss-Hermite parametrization of an LOSVD	18
1.13	The effect of different anisotropy structures on LOSVDs	23
1.14	Example of a surface of section	26
1.15	Illustration of over- and under-fitting in model selection	30

Zusammenfassung

Elliptische Zwerggalaxien (dE Galaxien) sind der häufigste Galaxientyp in den nahe gelegenen großen Galaxienhaufen und werden oft als die grundlegenden Bausteine von großen Galaxien betrachtet. Daher ist ihre innere Struktur von entscheidender Bedeutung für unser Verständnis der Strukturbildung und der Galaxienentwicklung. Die Masse und die Orbit-Struktur solcher Galaxien lassen sich mithilfe hoch entwickelter dynamischer Modellierungsverfahren ableiten. Aufgrund ihrer geringen Geschwindigkeitsdispersion und Flächenhelligkeit sind sie jedoch schwer zu untersuchen. Hochauflösende Spektroskopie und genaue, präzise Modelle sind erforderlich um ihre Form, die Verteilung ihrer dunklen Materie, ihre 3D-Orbit-Struktur, ihre stellaren Populationen und ihre anfängliche Massenfunktion zu untersuchen.

In dieser Arbeit wird eine Stichprobe von dEs des Virgo-Haufens mit Hilfe der Schwarzschild Modellierungstechnik untersucht, welche eine Galaxie als eine Überlagerung von Orbits modelliert. Während diese Technik bereits mehrfach zur Bestimmung von Galaxieneigenschaften wie etwa der Masse ihrer Schwarzen Löcher, ihres Masse-zu-Licht-Verhältnisses, und ihrer kinematischen Struktur verwendet wurde, war die Auswertung der Schwarzschild-Modelle nicht konsistent, da sie die variable Modellflexibilität nicht angemessen berücksichtigte. In dieser Arbeit wird ein neuartiger Bootstrap-Ansatz vorgestellt, der zum ersten Mal eine Schätzung der Flexibilität von Schwarzschild-Modellen ermöglicht, was deren Bewertung innerhalb eines neuen, verallgemeinerten Akaike Modellauswahl Ansatzes erlaubt. Der neue verallgemeinerte Ansatz ist vielseitig einsetzbar und nützlich im Allgemeinen für eine optimierte statistische Modellierung von ‘penalized’ Modellen. Im Kontext der Schwarzschild-Modellierung verbessert er den Mechanismus auf mehreren Ebenen. Er ermöglicht eine unvoreingenommene Bestimmung des besten Orbit-Modells, erlaubt die Optimierung der Model-Regularisierung, und verbessert die Bestimmung der LOSVDs (dem Input für die dynamischen Modelle).

Der neue Ansatz dient der sorgfältigen Modellierung der dE-Galaxien, mit dem Ziel ihre intrinsischen Eigenschaften zu ermitteln. Die wichtigsten Ergebnisse sind: die stellare Komponente von dEs ist räumlich homogen, da Masse-zu-Licht-Verhältnis, Alter, und Metallizität innerhalb des untersuchten Radialbereichs nur wenig variieren. Im Gegensatz dazu variieren diese Eigenschaften signifikant zwischen den verschiedenen dEs, was darauf schließen lässt, dass die dEs zu unterschiedlichen Zeiten mit der Sternbildung aufgehört haben. Im Vergleich zu normalen elliptischen Galaxien haben dEs einen unterdrückten Drehimpuls und eine Scheiben-ähnliche Orbitstruktur. Im Durchschnitt stimmt die stel-

lare Anfangsmassenfunktion von dEs weitgehend mit einer Kroupa Anfangsmassenfunktion überein, eine detaillierte Analyse zeigt jedoch, dass ihre Anfangsmassenfunktion mit dem Alter korreliert. Letzteres lässt sich durch eine Variation der Anfangsmassenfunktion mit der Entstehungs epoche oder durch unterschiedliche Grade einer zeitlich ausgedehnten Sternentstehungsgeschichte erklären. Die Dunkle-Materie Halos von dEs sind leicht ‘cored’ und haben eine runde Form, was auf einen milden Konflikt mit kosmologischen Standard-simulationen hindeutet, welcher möglicherweise durch eine angemessene Behandlung der baryonischen Rückkopplung in den Simulationen erklärt werden könnte. Die Korrelationen mit den Eigenschaften der Galaxie-Umgebung, des Drehimpulses und der Sternpopulation deuten darauf hin, dass die Halo Entstehung in dEs nicht universell ist und von den Umständen während ihres Gravitationskollapses abhängt. Alles in allem deuten die Ergebnisse darauf hin, dass die Stern-formenden Vorläufer von dEs unter extremen Bedingungen bei hohen Rotverschiebungen entstanden sind. german

Abstract

Dwarf Elliptical (dE) galaxies are the most abundant type of galaxies in the nearby large galaxy clusters and often considered as the fundamental building blocks of giant galaxies. As such their intrinsic structure is crucial to our understanding of structure formation and galaxy evolution. The mass and orbit structure of such galaxies can be inferred using sophisticated dynamical modelling techniques. However, their low velocity dispersions and surface brightness makes them difficult to study. High resolution spectroscopy and accurate and precise models are required to study their shapes, dark matter distribution, 3D orbit structure, stellar populations, and initial mass function.

This thesis sets out to study a sample of dEs located in the Virgo cluster by the use of the Schwarzschild modelling technique, which models galaxies as a superposition of orbits. While the technique has been used numerous times to constrain galaxy properties like black hole masses, mass-to-light ratios and kinematic structure, the evaluation of the Schwarzschild models was not always consistent and accurate since it lacked a proper accounting of the variable model flexibility. This thesis introduces a novel bootstrap approach that, for the first time, enables an estimation of the flexibility of Schwarzschild models which allows their evaluation within a new generalized Akaike model selection framework. The new generalized approach is versatile and useful for the optimized statistical modelling of penalized models in general. In the context of Schwarzschild modelling it improves the machinery on several levels: it allows an unbiased determination of the best orbit model, it permits the optimization of the model regularization, and it improves the recovery of line-of-sight velocity distributions (the input for the dynamical models).

The new approach is adopted to carefully model the dE sample with the goal to recover their intrinsic properties. The main findings are: The stellar component of dEs is spatially homogeneous, as mass-to-light ratios, age and metallicity vary little within the investigated radial range. Conversely, these properties vary significantly between different dEs which suggests that the dEs were quenched in different epochs. Compared to normal Ellipticals dEs have a suppressed angular momentum and a disc-like orbit structure. On average, the stellar initial mass function (IMF) of dEs are broadly consistent with a Kroupa IMF, but a detailed analysis shows that their IMF is correlated with the age. This can be explained by a variation of the IMF with formation epoch or varying degrees of extended star formation history. The dark matter halos of dEs are mildly cored and have round shapes, suggesting a mild tension with standard cosmological simulations, which could still be explained by proper treatment of baryonic feedback in simulations. Correlations with environment,

angular momentum and stellar population properties suggests that the halo formation in dEs is not universal but depends on the circumstances during the gravitational collapse phase. All together the results suggest that the star-forming progenitors of dEs are not comparable to local star-forming dwarfs as the halos of dEs have formed in more extreme conditions at high redshifts.

Chapter 1

Introduction

A mere 101 years ago, in 1923, our understanding of the Universe exploded. Or at least humanity's understanding of its size. Edwin Hubble identified and measured the periods of Cepheid variables in the Andromeda Nebula (M31), which allowed him to measure their absolute brightness and, consequently, their distance to Earth. The result, several hundred kiloparsecs, settled the so-called 'Great Debate' held a few years prior. It confirmed that our Milky Way galaxy was not the entirety of the Universe but just one of many galaxies in the observable Universe. In the following century, countless such 'external' galaxies of various shapes and sizes were discovered, establishing a huge diversity of galaxy morphology which required classification and physical explanation.

By far the most well-known classification scheme is the 'Hubble tuning fork' (Hubble, 1927, 1936) illustrated in Fig. 1.1. It categorizes galaxies along a sequence according to their appearance to the eye. Galaxies that have a smooth and featureless 3-dimensional ellipsoidal structure, so-called Ellipticals (E), are located on one end of the fork. Galaxies with a seemingly more complex substructure, so-called Spirals (S), populate the other end of the Hubble fork. Spirals consist of a very flat stellar disc and often have spiral-arms and central bulges embedded within said disc.

The Elliptical galaxies are sequenced according their flattening as they appear on the sky. A suffix 'n' following after the 'E' indicates a galaxy's ellipticity. The suffix is defined as $n = 10 \cdot \epsilon = 10 \cdot (1 - b/a)$ where a is the semi-major and b the semi-minor axis of elliptical isophotes that approximate the galaxy's brightness profile. A spherical galaxy will be denoted 'E0', while the flattest Ellipticals (found to have $\epsilon \approx 0.7$) are classified as 'E7'.

The Spirals make up the 'fork crown' of the tuning fork by forming two parallel sequences. If the arms of a spiral galaxy are attached to a strong central bar it is categorized as a 'SB'. If a spiral galaxy has no prominent bar it is categorized as a normal (or ordinary) 'S'. While the fraction of Spirals with bars may vary strongly with various properties such as total mass or redshift, broadly speaking about 30% to 60% of Spirals are barred (e.g. Erwin, 2018). Both of these spiral branches are further partitioned with suffixes 'a', 'b', and 'c' depending on the prominence of their central bulge component and how closely wound and defined their spiral arms are.

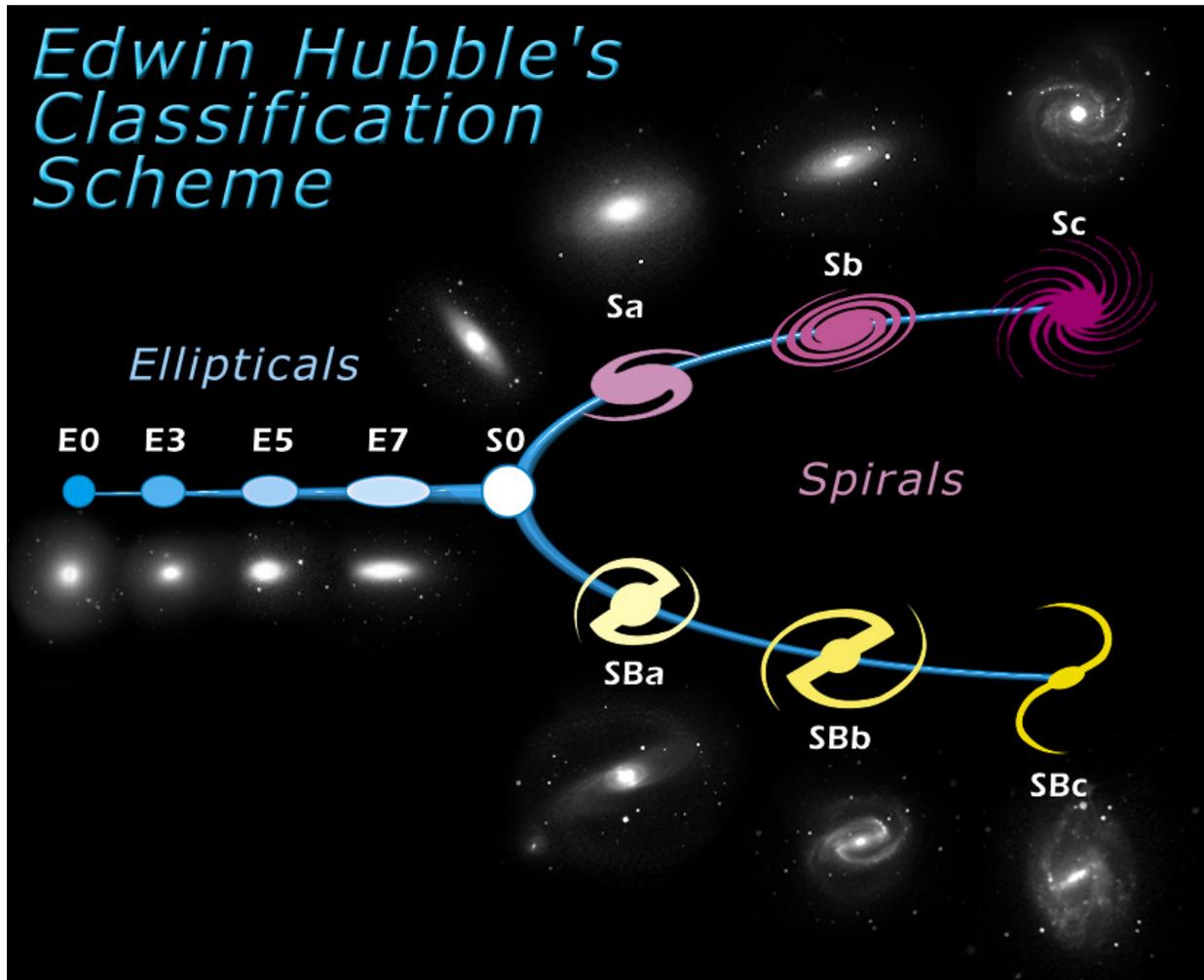


Figure 1.1: The classical Hubble tuning fork diagram classifying galaxies into Ellipticals (E) and Spirals which occur barred (SB) and unbarred (S). Many galaxies are Irregular (Irr) and can not be properly placed in this classification scheme. Image credits: <https://esahubble.org/images/heic9902o/>.

At the juncture between the Ellipticals and Spiral sequences are the lenticular galaxies (or ‘S0’ galaxies). Like the spirals, they have a disc component, but without embedded spiral arms. Instead, their stellar distributions are smoother, like that of the Ellipticals, and they have a more prominent bulge component than Spirals.

1.1 The ETG-LTG dichotomy

The Hubble classification is often ambiguous, it neglects orientation effects which can make the same galaxy appear very different depending on the viewing angles, and numerous galaxies, so-called Irregular galaxies (‘Irr’), can not easily be placed in the Hubble cate-

gories at all. The scheme is, first and foremost, a morphological one and not a physically motivated one.

However, as observational capabilities increased in the decades after Hubble, several physical properties were found to align with the Hubble ranking, making the classification scheme and its more advanced modifications (Kormendy & Bender, 1996; Faber et al., 1997) useful to this day. For example, investigations of the integrated colour of galaxies (de Vaucouleurs, 1961) revealed that Ellipticals are redder than Spirals. The colour of a galaxy is associated with the age of a galaxy's average stellar population, as bluer stars have a shorter lifetime. This makes the galaxies integrated colour redder over time, unless new (blue) stars are constantly being formed.

This difference in age and colour is also why one could be misled to believe that the Hubble tuning fork is an evolutionary sequence from left to right. This is the reason why Ellipticals and S0 galaxies are commonly grouped as so-called Early-Type galaxies (ETG) which have red stellar populations that have ceased star formation some time ago, whereas Spirals and Irregulars which are bluer and still in the process of forming new blue stars are grouped as Late-Type galaxies (LTG). In reality the picture is much more complex and the Hubble fork is far from a strict evolutionary sequence going from left to right. Still, the ETG/LTG naming convention is convenient to this day because it turns out that the two galaxy classes distinctly differ in many of their *intrinsic* physical properties.

ETGs appear old and red because they have stopped active star formation, as they lack a sufficiently large cold gas reservoir from which new stars could be formed. This lack of gas is not necessarily because they have fully consumed all their available gas, but more likely due to a so-called *quenching* process. Such a process can remove the gas from a galaxy, or at least make it too hot to form new stars. Many different quenching mechanisms have been proposed and it is likely that the dominant mechanism changes depending on the total galaxy mass and environment. LTGs on the other hand still have large cold gas and dust reservoirs that are mostly confined in their disc. This allows them to form new stars to this day. As a result of this ongoing star formation, their interstellar medium also tends to be fairly metal-enriched.

The ETG/LTG dichotomy is not confined to the properties of their stellar populations. Photometrically, the light distribution of ETGs usually follow steeper central surface brightness profiles approximately described by the $R^{\frac{1}{4}}$ law (de Vaucouleurs, 1948), whereas the discs of LTGs follow exponential brightness profiles (see details later, eq. 1.1). The kinematic structure of ETGs is predominantly supported by their large stellar velocity dispersion (i.e. random motion of the stars), whereas the kinematics of LTGs is more ordered (i.e. net rotation of stars) with a high angular momentum orientated perpendicular to their disc. ETGs and LTGs are also known to follow a strong environment-morphology dichotomy (Dressler, 1980; Binggeli et al., 1987; Geha et al., 2012). While the ETGs tend to populate the denser environments of clusters, LTGs are usually more isolated. This suggests that the environment has played a crucial role in whether a galaxy's star formation was quenched or not.

This thesis is concerned with a subspecies of ETGs: the dwarf ellipticals (dE) which sometimes are also referred to under the umbrella term 'spheroidals'. Even though the

name suggests that dEs are simply dwarf versions of the brighter Ellipticals (in the following ‘ordinary’ E) their surface brightness distributions are very distinct to them (Sec. 1.2.1) and instead alludes to a much closer link of dEs to LTG dwarfs (Sec. 1.2.2). In this sense, dEs occupy an interesting ‘intermediate’ spot between the classical ETG/LTG distinction. Studying the intrinsic structure of these ‘transitional’ objects could aid in our understanding of the mechanisms responsible for the large morphological diversity of extragalactic galaxies as a whole.

However, up until now, the faint surface brightness and low velocity dispersions of these galaxies have prevented studies of their detailed intrinsic structure. Instead, existing evidence is limited to ‘surface-level’ constraints on their observables (Sec. 1.3) and even those were often biased due to the poor resolution of the instruments that were used to observe them (Chapter 3). This thesis sets out to be the first dE study that is able to employ sophisticated dynamical modelling (Sec. 1.4) which allow the recovery of the 3D mass (stars, dark matter, black holes) and orbit structure. This is made possible by obtaining data with unprecedented spectral resolution and by improving the dynamical modelling technique itself (Chapter 2) by embedding it in a powerful model selection paradigm (Sec. 1.5). The detailed and comprehensive information about the stellar and dark matter components obtained in this way allows the inference of the stellar initial mass function (IMF) and its condition-dependent variability (Chapter 3), and points towards a distinct formation path of dEs as remnants of high redshift late-type dwarfs (Chapter 4). It is less their recent evolution that made the dEs appear the way they do today, but more so the initial conditions and environment at the time they have assembled in.

1.2 Dwarf ellipticals

1.2.1 Dwarf ellipticals - A distinct subclass in the ETG sequence?

The structure of massive ETGs with stellar masses $\log(M_*/M_\odot) \geq 10$ have been studied extensively with dynamical models (e.g. Gebhardt et al., 2003; Thomas et al., 2007; Auger et al., 2009; Thomas et al., 2011; Cappellari et al., 2011; Rusli et al., 2013; Neureiter et al., 2023; Mehrgan et al., 2024). In comparison, little is known about the intrinsic structure of dEs despite the fact that they are by far the most abundant type of galaxy in the nearby galaxy clusters. In parts, this lack of dynamical studies is because they are fainter and have very low velocity dispersions that require higher resolution spectroscopy to be measured correctly.

There exists no ubiquitously accepted naming convention among dwarf galaxies, but broadly speaking, dE-like galaxies are quiescent galaxies with total B-band magnitudes fainter than -18 mag that have a relatively low central surface brightness (SB). This distinction is made because the majority of ETGs below this magnitude tend to have shallower surface brightness profiles than the more luminous ETGs¹ which was first noticed by Baade (1944).

¹Though there are exceptions, like M32, which have steeper SB profile akin to the more massive ETGs.

Among the dE-like galaxies, several subtypes and conventions exist in the literature. For example, some authors differentiate between the even fainter dwarf Spheroidals (dSphs) with $M_B \geq -14$ mag and brighter dEs (Gallagher & Wyse, 1994; van den Bergh, 2000; Seo & Ann, 2022), while others purposely group them together, simply referring to them as ‘spheroidals’ (Ferguson & Binggeli, 1994; Kormendy & Bender, 2012; Zöller et al., 2024). Since this thesis is concerned with a sample of relatively bright dwarfs ($M_B \approx -17$ mag) this thesis sticks to the stricter sense of the definition unless stated otherwise. This means dEs are the *more luminous* ‘spheroidals’ with absolute magnitudes between -18 mag and -14 mag.

With a few exceptions (NGC 147, NGC 185, NGC 205) the majority of nearby dEs, if defined as above, is not found within the local group, but in clusters like Virgo and Fornax. Conversely, the majority of the fainter dSphs, at least those one is able to study with current tools, are satellites of the massive local group galaxies (Milky Way and Andromeda).

As mentioned above, the distinct nature of dEs was first noticed in their surface brightness and effective size. This is quantified and illustrated in Fig. 1.2. Ordinary Ellipticals follow a tight scaling correlation: the Fundamental plane (Dressler et al., 1987; Djorgovski & Davis, 1987) which relates their central velocity dispersion σ_o , effective radius r_e , and mean surface brightness SB_e to one another. If two of these parameters are known, the other can be predicted using this relation. When projecting the fundamental plane to the SB_e - r_e subspace, it becomes the Kormendy-relation (Kormendy, 1977). This relation is evident in the top panel of Fig. 1.2. As the effective radius increases, the ‘ordinary’ Ellipticals (grey) become progressively fainter. Conversely, the ‘spheroidals’ (i.e. dEs, UDGs, dSphs, dE) all fall outside the Kormendy relation as their surface brightness is much fainter and their size much bigger at the same total Luminosity.

This disparity is clear when scaling SB_e and r_e versus the total absolute magnitude M_{tot} (bottom two panels of Fig. 1.2). The higher the total luminosity of an ‘ordinary’ elliptical, the fainter its surface brightness and larger its size, while the surface brightness of the dwarfs tend to follow the opposite trend. In the small magnitude range $[-16\text{mag}, -18\text{mag}]$ where both ‘ordinary’ ellipticals and dEs coexist, the two types are clearly disjoint with dEs being significantly fainter and more extended for the same total luminosity.

This difference in the mean surface brightness profiles of ‘ordinary’ ellipticals and dEs is connected to a difference in the shape of their light profiles. This can be measured by measuring a galaxy’s *Sersic index* n , which is a measure of the curvature of its surface brightness profile. The surface brightness profile $\mu(r)$ is obtained by fitting elliptical isophotes to the 2D galaxy image. Then the Sersic index is obtained by fitting the profile with a Sersic function (Sérsic, 1963) which is defined as:

$$SB(r) = SB_e \cdot \exp\left(-b_n \left(\left(\frac{r}{r_e}\right)^{1/n} - 1\right)\right) \quad (1.1)$$

where SB_e is the surface brightness at the effective radius r_e and b_n is a normalization factor which depends only on n . A low Sersic index n indicates a shallow profile, conversely a high index indicates a steep profile. For $n = 4$ this equals the $R^{\frac{1}{4}}$ law (de Vaucouleurs, 1948) and for $n = 1$ the profile becomes an exponential one.

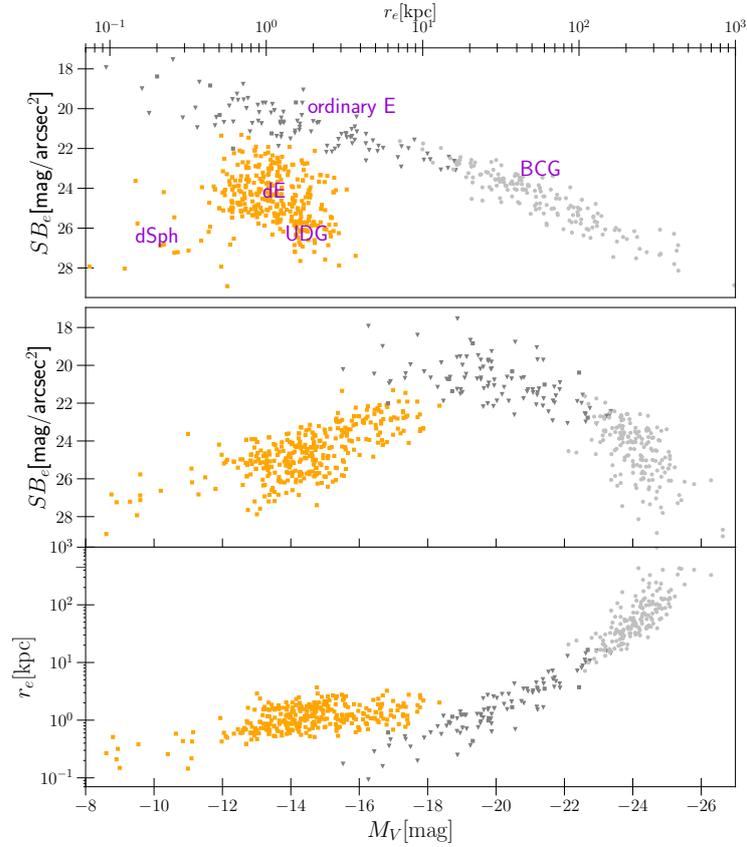


Figure 1.2: Correlations of the mean surface brightness SB_e at the effective radius versus total magnitude M_{tot} and the effective radius r_e for various ETG types. The data is adopted from Zöller et al. (2024). Here, dEs are grouped together with dSphs and ultra-diffuse galaxies (UDGs) under the umbrella term ‘spheroidals’(orange). UDGs are sometimes considered a separate galaxy class, as they are even fainter and more extended than the dEs. dSphs occupy the low luminosity branch of the spheroidals. Ordinary ETGs follow the Kormendy relation and are more concentrated at the same total luminosity. The brightest cluster galaxies (BCGs) form the upper end of the massive ETG sequence and can be considered its own ETG subclass (e.g. Kluge et al., 2020).

‘Ordinary’ ellipticals usually have high Sersic indices² in the range of $n \approx 3 - 10$, while dEs have shallower profiles with $n \approx 1 - 2$ (e.g. Binggeli & Jerjen, 1998; Janz et al., 2014).

This is illustrated in Fig. 1.3 which shows the Sersic indices of various ETGs types vs absolute magnitude. While the Sersic indices of the early-type spheroidals (dEs, UDGs,

²The *most* massive ETG actually differ again as they are not well-fitted by a single Sersic profile since they have cored profiles. Instead, a ‘core-Sersic’ profile is a better approximation for these galaxies (Graham et al., 2003). This is one of the features of the well-studied ‘E-E dichotomy’ in the population of ‘ordinary’ ellipticals (e.g. Kormendy et al., 2009). dEs and spheroidals do not belong to either of these classes and should be thought of as an entirely different 3rd class of ETGs.

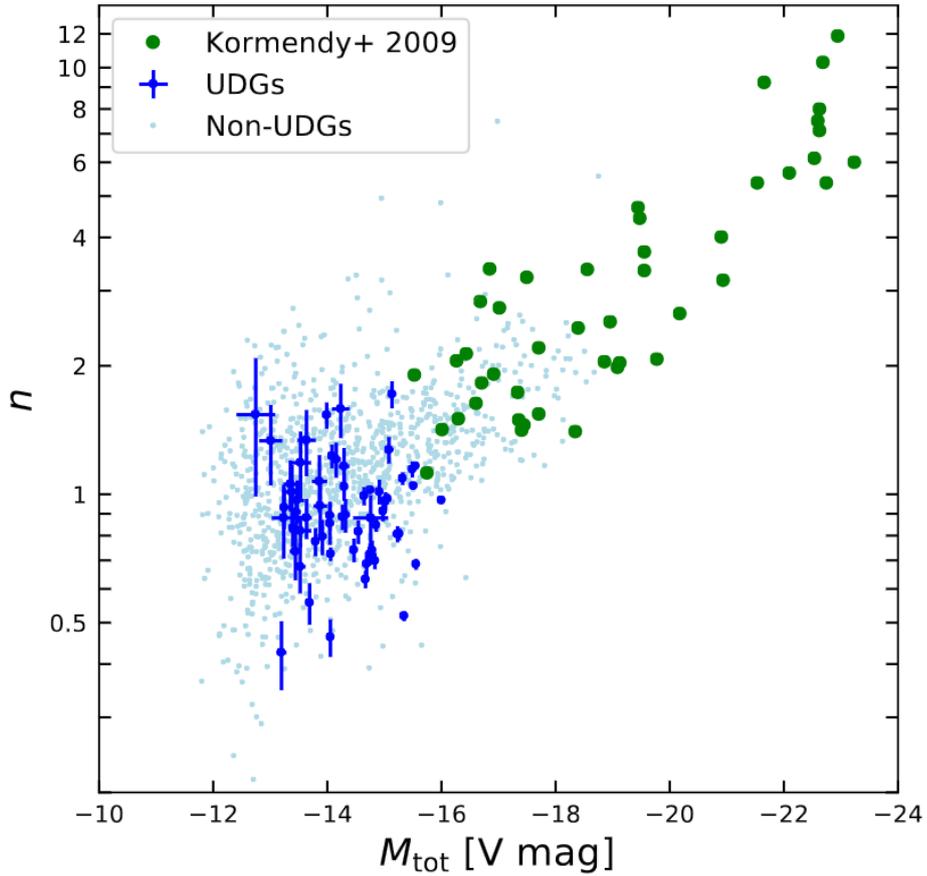


Figure 1.3: Sersic index n vs absolute magnitude (V-band) for early-type galaxies. ‘Non-UDGs’ include the dE and dSph galaxies. In *green* ‘ordinary’ ellipticals are shown. Early type galaxies fainter than ~ -16 mag are on average consistent with an exponential light profile. Image credits: Zöller et al. (2024).

dSphs) do not appear to be particularly disjoint from the sequence of classical ellipticals (green), it is still notable that almost all dwarfs below $M_B > -18$ mag have very shallow light profiles. In fact, they are nearly indistinguishable from the surface brightness profiles of spiral galaxies, which are generally well-fitted by an exponential profile ($n = 1$).

Further indication that dEs are distinct from ‘ordinary’ ellipticals comes in the form of their Luminosity function. The Luminosity function predicts the number of objects within some volume as a function of total Luminosity or mass of these objects. Luminosity functions of galaxy populations are typically well-fitted by a Schechter fit model (Schechter, 1976).

The Luminosity functions of dEs and ‘ordinary’ Es, as analysed by Sandage et al. (1985), are shown in Fig. 1.4. They distinguish two classes based on the steepness of their surface brightness profiles. Individually, the two galaxy classes are fitted well by a Schechter model, but the number counts of the combined population (E + dE) differs

significantly from a Schechter model. Furthermore, the individual Luminosity functions of Es and dEs follow very different Schechter distributions and the two populations only overlap for a small magnitude range $M_{\text{tot}} \in [-18, -16]$ mag. In other words, there exists essentially no very small Es or, vice versa, big dEs.

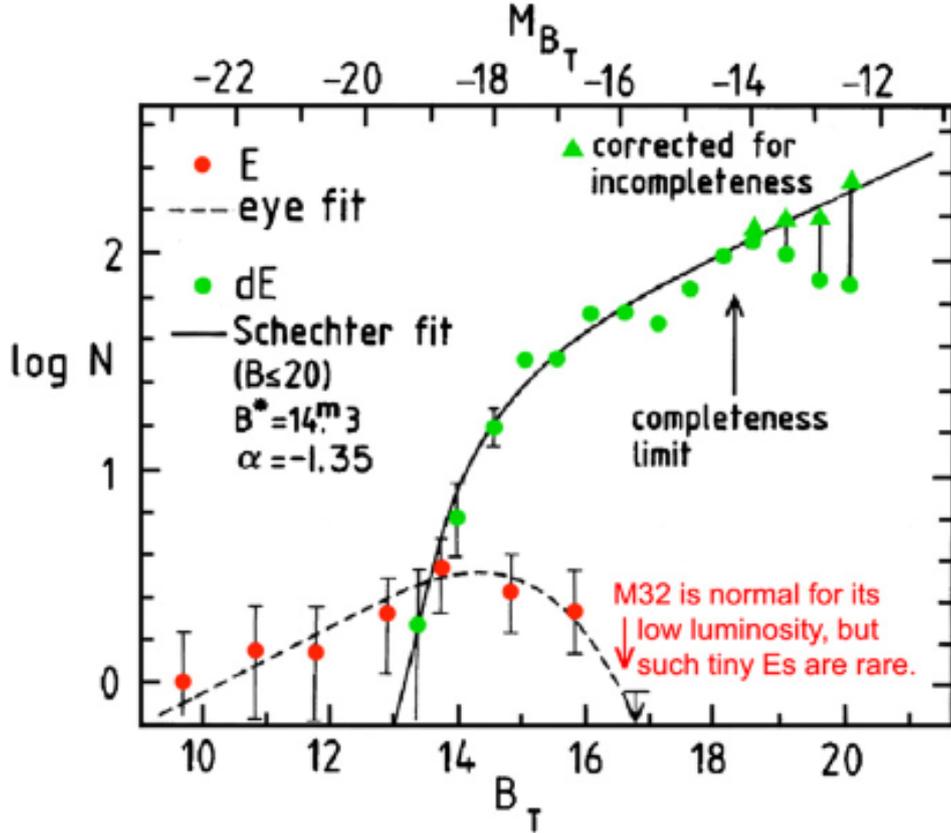


Figure 1.4: The Luminosity functions of E and dE galaxies as shown in Kormendy & Bender (2012), who adopted it from Sandage et al. (1985). dEs and E were separated based on the steepness of their surface brightness profile, i.e. essentially their Sersic index. Separately, the E and dE populations are fitted well by a Schechter fit model (Schechter, 1976), but the combined population would not be fitted well by such a model.

1.2.2 Dwarf ellipticals - Remnants of spiral galaxies?

The disparity in their luminosity functions suggests that ‘ordinary’ Es and dEs must have formed via different evolutionary channels. Since ‘ordinary’ ellipticals are believed to form hierarchically from (successive) major mergers (e.g. Hopkins et al., 2008) a different scenario would be needed for dEs and Sphs. The commonly proposed scenario is that the dEs are the *remnants* of quenched late-type galaxies (specifically loose dwarf spirals Sc and irregulars

Irr). This scenario is supported by several observed similarities that dEs have with those late-type galaxies.

As mentioned above, the Sersic indices of dEs are nearly as shallow as that of exponential disc profiles ($n = 1$). This implies they follow similar light profiles, which is reflected in their surface brightness scaling relations. This is illustrated in Fig. 1.5 which is the analog of Fig. 1.2 but includes late-type galaxies. The late-type galaxies (blue) follow the same trends as the dEs/spheroidals (green) and their distributions overlap such that they would be indistinguishable if they were not colour-coded in the Figure.

The similarity of dE and Sc-Im galaxies goes well beyond just the overall shape and surface brightness profile. When studied in detail, many dEs/Sphs are found to contain photometric features that are typical for LTGs such as faint spiral arms, bars and discs (Jerjen et al., 2000; Geha et al., 2003; Lisker et al., 2006a; Kormendy & Bender, 2012; Penny et al., 2014). An example of this is shown in Fig. 1.6 which shows a dE that appears smooth and featureless at first glance but after removing the elliptical light profile face-on spiral arms are revealed.

Similarly, the presence of such faint discs can also be detected in some edge-on dEs, like VCC 2048 (Fig. 1.7). Edge-on discs embedded in the dEs lead to distortions in elliptical isophotes, making them appear more *discy*. This deviation from a perfect elliptical isophote can be quantified using the fitting technique of Bender & Moellenhoff (1987) which quantifies these deviations using a Fourier expansion. When the even Fourier coefficients (usually termed a_4 and a_6) are positive, the isophotes are more discy, when they are negative, the isophotes are more boxy. Both cases are illustrated in Fig. 1.8. Positive a_4 have been measured in many dEs indicating the presence of embedded discs in them (e.g. Kormendy & Bender, 2012).

Beyond these small-scale photometric features, a significant portion of dEs also exhibit blue centres that hint at recent star formation and show traces of dust and gas (Lisker et al., 2006b) similar to late-type dwarfs. Investigations of the star-formation histories of dEs (and dSphs) suggest that they have build up their stars over an extended period of time via periodic bursts, spread out over several Gyrs (e.g. Weisz et al., 2011; Seo & Ann, 2023; Romero-Gómez et al., 2024). This suggests dEs have had gas available for prolonged periods of time, making them indistinguishable from late-type dwarfs during these times, until they were eventually quenched completely.

Further evidence of a connection of dEs to dwarf LTGs could come from a similarity in their observed rotation signals. However, unlike the photometric results, the kinematic similarities are far less conclusive. Some studies find significant rotation in dEs statistically indistinguishable from that of late-type dwarfs (e.g. Simien & Prugniel, 2002; van Zee et al., 2004), while others find surprisingly small rotation (e.g. Bender et al., 1991; Geha et al., 2003). Therefore it could be that a significant portion of the dE population is ‘pressure-supported’, which, in the context of galaxy dynamics, means their flattening is not due to ordered stellar motion (net rotation) but due to an anisotropic, unordered motion (velocity dispersion).

All together, it appears that the distinction between dEs and Irrs is not always unequivocal and that there likely is a fluent transition between the two. This similarity between

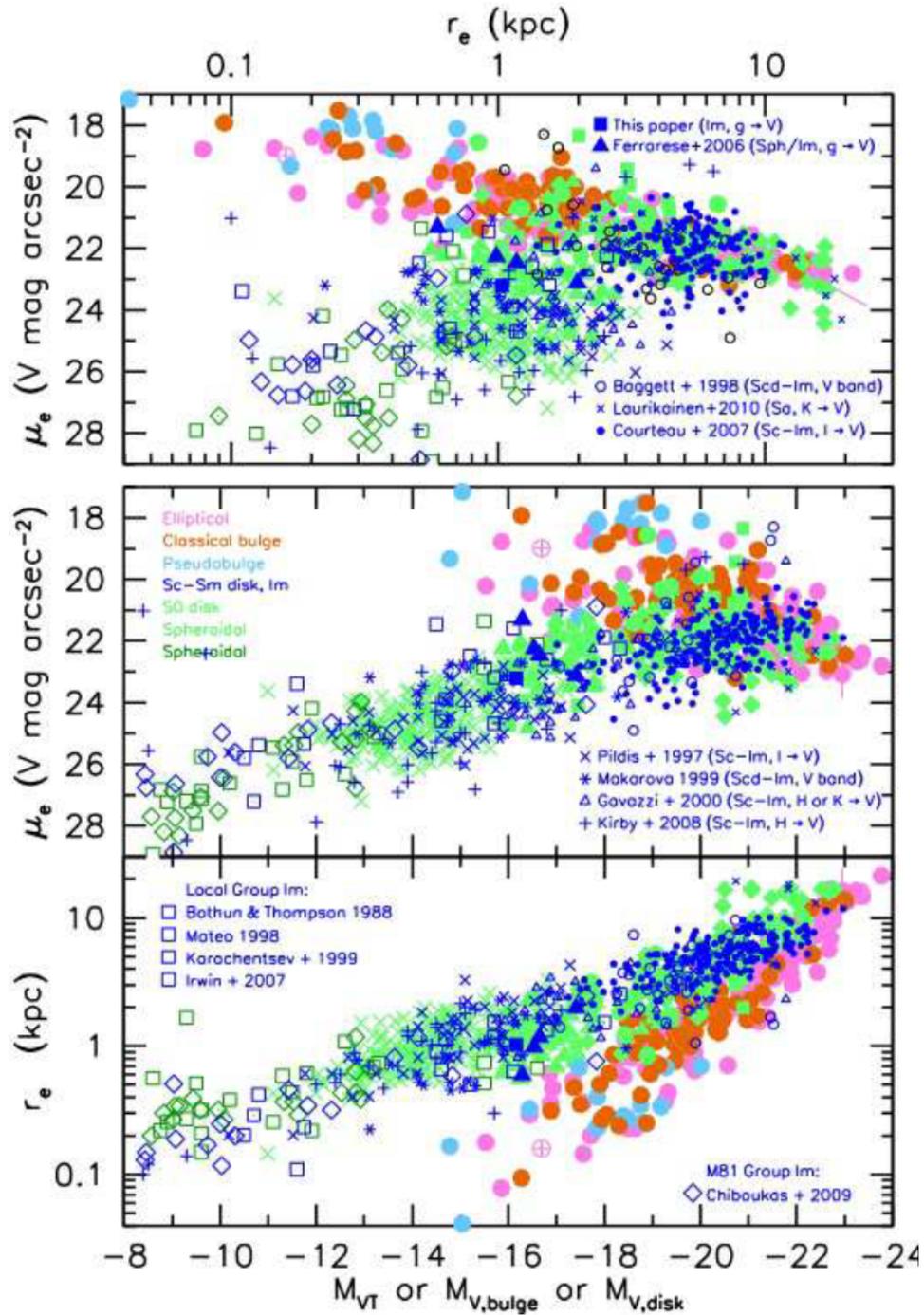


Figure 1.5: Similar to Fig. 1.2, but including LTGs (*blue*). Again ‘spheroidals’ (*green*) is used in the broader sense, which means dEs are included. Photometrically, dEs/spheroidals behave very similar to late-type dwarfs and are clearly disjoint from the population of ‘ordinary’ Ellipticals. Image credits: Kormendy & Bender (2012).

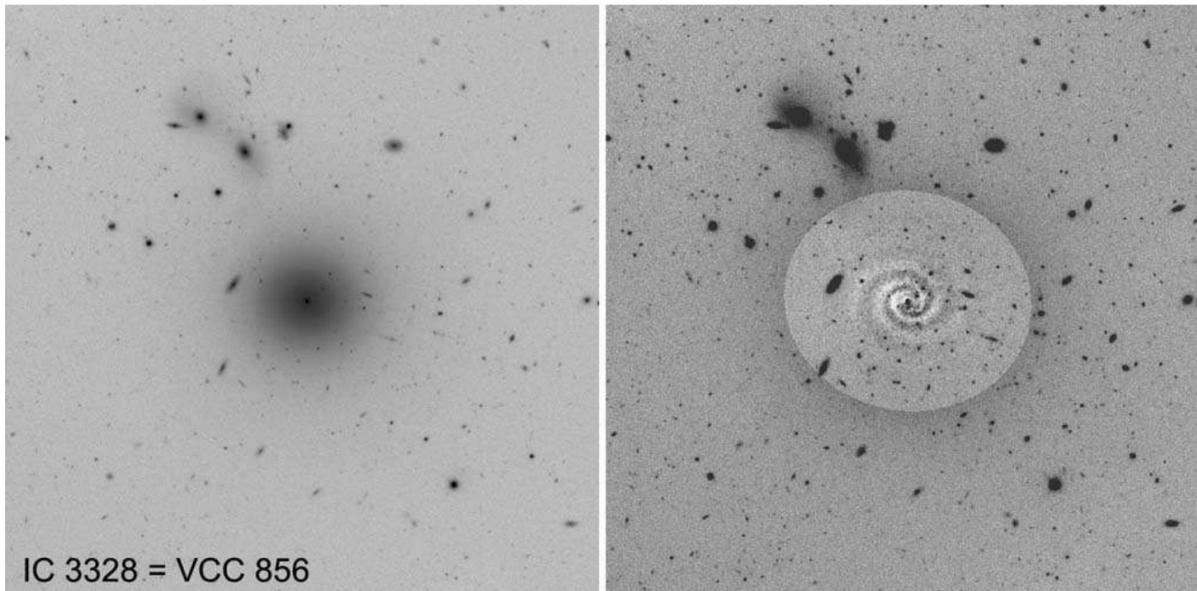


Figure 1.6: *Left:* R-band image of a dE (VCC 856). *Right:* The residual image after the elliptical component is removed from the light. It reveals that this dE harbours face-on spiral arms. Image credits: (Jerjen et al., 2000).

the two dwarf types has led Kormendy & Bender (2012) to place them in a revised version of the physically motivated galaxy classification scheme of van den Bergh (1976). An illustration of this alternative to the Hubble fork is displayed in Fig. 1.9. Instead of placing S0 galaxies at the junction between ellipticals and spirals, the S0's form a separate arm that is parallel to the sequence of spirals (barred and normal spirals are grouped together). For both arms the bulge-to-total light ratios decreases from left to right, and a galaxy is attributed to either the S0 or the spirals arm depending on its gas content and star formation activity. Kormendy & Bender (2012) argue the dEs/Sphs form a continuous sequence with the S0 galaxies, and they are the ‘bulgeless’ extension of the S0s.

The physical interpretation of this scheme is that some quenching process has transformed the spirals to become the counterpart in the ‘red and dead’ S0 arm (e.g. an Sb can be transformed to a S0b). Several transformation processes that have facilitated this transformation have been proposed. For example, while progenitor Irr's have entered larger galaxy clusters, their interaction with the intra-cluster medium could have rapidly stripped them of their remaining gas via a process called ram-pressure-stripping (Gunn & Gott, 1972). Similarly, tidal interaction with larger cluster or group members (Moore et al., 1999) could have suppressed the star formation of Irr's turning them into the dEs observed in the present. Both these scenarios are strongly supported by the environment-morphology dichotomy, which is the observation that the majority of dEs are found within the dense galaxy clusters whereas Irr avoid dense environments, but dominate the dwarf population in the field (Binggeli et al., 1987; Geha et al., 2012).

Alternative to these *externally* induced quenching processes an *internal* mechanism

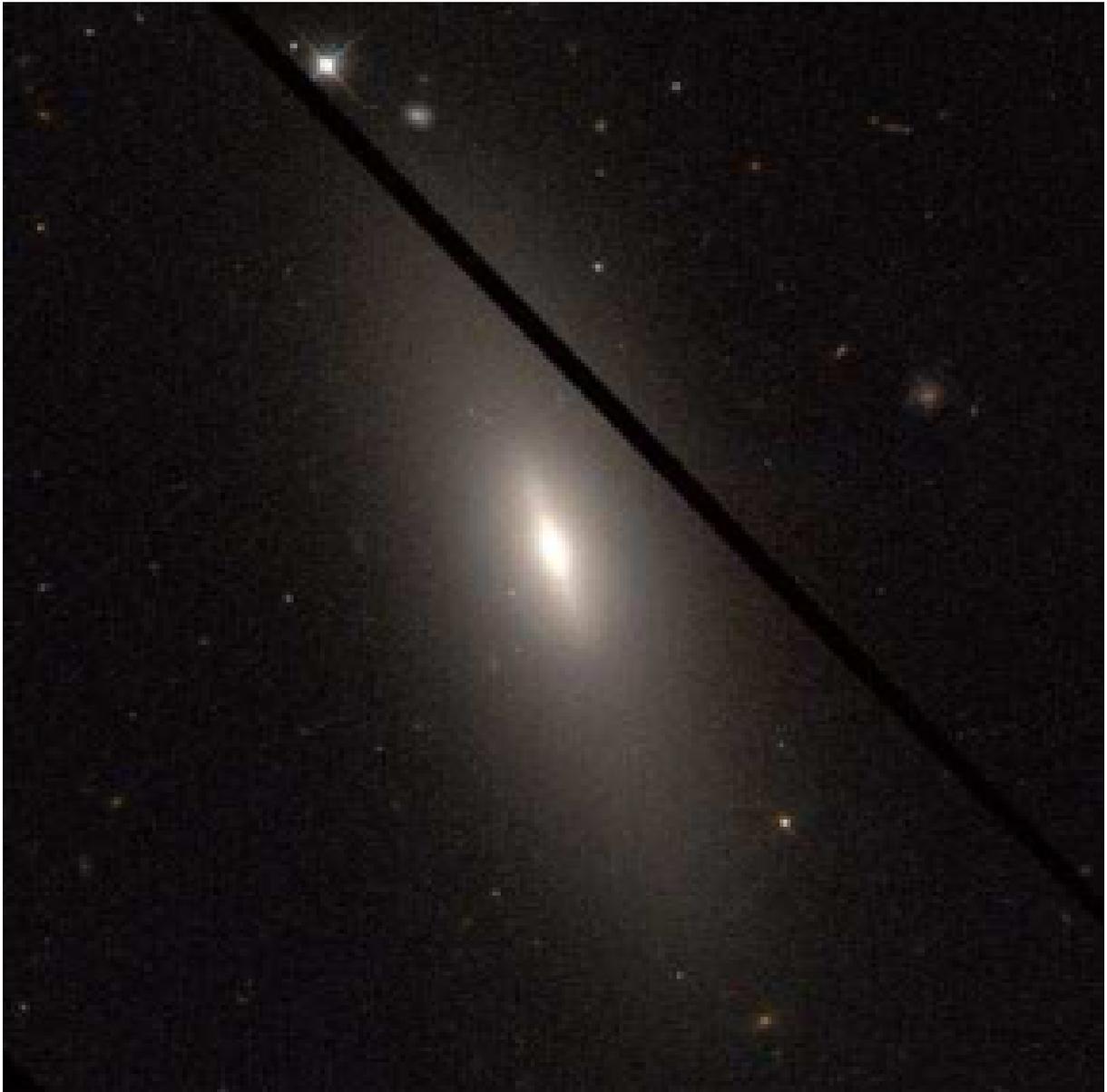


Figure 1.7: Colour image of the dE VCC2048 from Kormendy & Bender (2012) which was obtained from HST ACS images in g and z band. The galaxy contains a faint, near edge-on, disc which leads to deviations of its isophotes from a pure elliptical form.

could also be responsible for the gas loss dEs must have suffered. The potential well of less massive galaxies is naturally smaller than those of more massive ones, therefore dEs may lose their gas more easily than ‘ordinary’ Es due to feedback from active galactic nuclei, supernovae, or stellar winds (e.g. Dekel & Silk, 1986). While dEs/dSphs may have lost a lot of their gas due to a more turbulent history, spirals have had a more gradual and slow build up such that they were able to retain most of their gas.

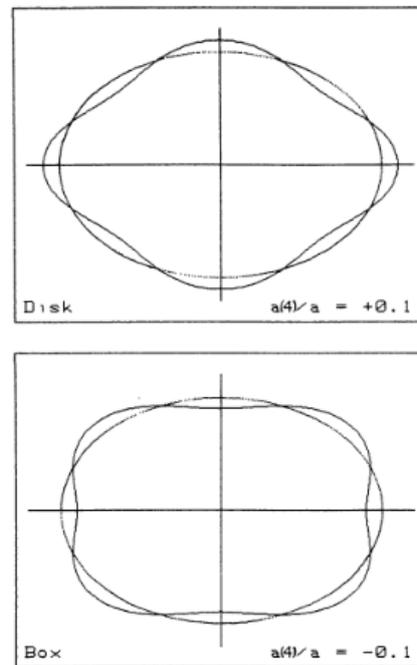


Figure 1.8: An illustration how a discy (top) and a boxy (bottom) isophote deviates from a perfect ellipse with semimajor axis a . The discy isophote in the top panel has $a_4/a = +0.1$. The boxy isophote in the bottom panel has $a_4/a = -0.1$. Image credits: Bender et al. (1988).

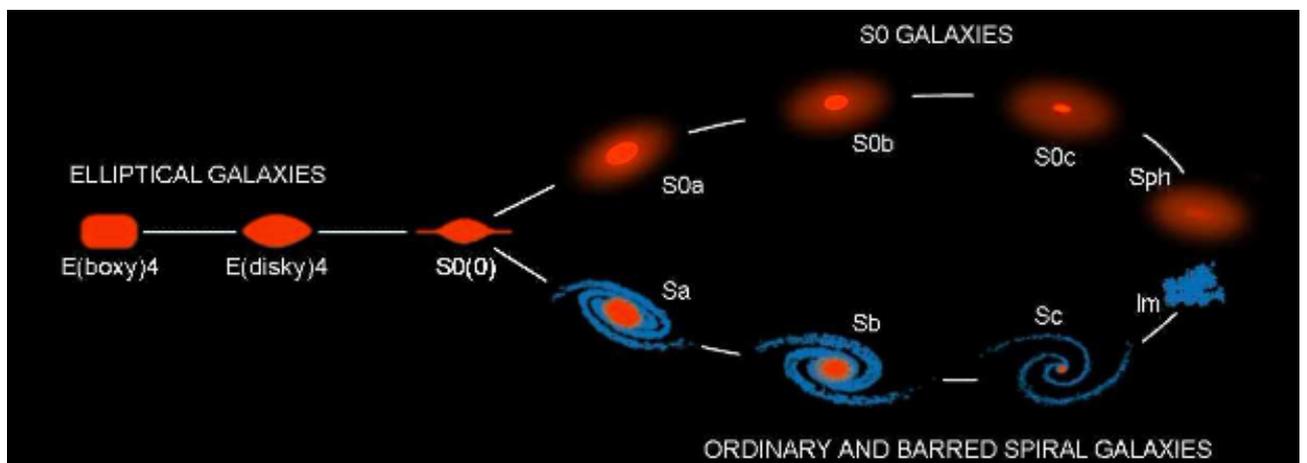


Figure 1.9: The morphological parallel-sequence of galaxies from Kormendy & Bender (2012) which is a revised version of the scheme by van den Bergh (1976). The definition of spheroidal galaxies (Sphs) includes the dEs discussed in this thesis. In this classification scheme, dEs are the quiescent counterparts of the star-forming Im/Irr galaxies.

The above examples of plausible mechanisms is far from comprehensive, further examples include starvation (e.g. Larson et al., 1980), reionization feedback (Bullock et al., 2000), dwarf mergers (e.g. Cloet-Osselaer et al., 2014), and tidal debris (Barnes & Hernquist, 1992). One of the goals of this thesis is to investigate which of these mechanisms are the most likely ‘culprits’ that have led to the transformation of late-type progenitors. Connected to this is the question of how dEs and dwarf LTGs can coexist in the local Universe despite their similarities. The mechanism responsible for the cessation of star formation has to be discriminatory: halting star formation in some, but not all dwarf galaxies.

By analysing the intrinsic structure, dark matter distributions and orbit occupations of these galaxies one may gain deeper insight into what has stopped the star formation in some dwarf galaxies. Information about the intrinsic structure is extracted by constructing dynamical models of these galaxies. And this thesis presents the first attempt of this undertaking for dE galaxies, which makes use of a technique (‘Schwarzschild modelling’, Schwarzschild, 1979) that is sufficiently robust and accurate to recover the intrinsic structure in an unbiased manner. A total of 9 dEs, all located in the Virgo cluster, are modelled using this technique. An image of the Virgo cluster with the 9 dEs investigated in this thesis is shown in Fig. 1.10. Dynamical modelling in general, and the Schwarzschild technique specifically, is outlined in Sec. 1.4, the data preparation and input required for these models is discussed in the following section.

1.3 The observables

To study the intrinsic structure of dEs dynamical modelling techniques, like they are discussed in Sec. 1.4, are indispensable. The machinery employed in this thesis requires a *preparation* of the available data sets from imaging and spectroscopy such that it can be used as input by the models. This preparation is briefly outlined in the following.

1.3.1 Photometry

Advanced models of galaxies are usually comprised of several mass components such as dark matter, stars, or black holes. While the dark mass components can only be inferred dynamically, imaging data provide a direct boundary constraint on the luminosity distribution (but not the luminous mass). One can implement this boundary constraint by describing the luminous mass density as $\rho_* = \Upsilon \cdot \nu$ where Υ is the stellar mass-to-light ratio and ν the 3D luminosity deprojection.

This requires a *deprojection* of the 2D surface brightness $I(x', y')$ data, i.e. the inverse of:

$$I(x', y') = \int_{-\infty}^{\infty} \nu(x, y, z) dz \quad (1.2)$$

where the intrinsic coordinate system is defined such that z is orientated parallel to the line-of-sight. Generally this inversion is only unique if the galaxy is assumed to follow very strict symmetry assumptions (e.g. if it is spherical). However, even for less strict



Figure 1.10: Image of the Virgo cluster obtained from the Sloan Digital Sky Survey (Blanton et al., 2017). Marked in red are the location of the dEs analysed in this thesis. Green shows the bright cluster galaxy M87 which is approximately in the cluster centre.

symmetry assumptions, like axisymmetric and triaxial galaxies, the allowed solution space is restricted considerably. Under such symmetry assumptions, deprojection algorithms exist that allow one to probe the available solution space (Magorrian, 1999; Cappellari, 2002; de Nicola et al., 2020).

In the *axisymmetric* case the 3D luminosity distribution ν is described fully by its description in the 2D meridional plane (R, z) and a single viewing angle: the inclination i . However, even if i is known, an inversion of eq. 1.2 is only unique if the meridional plane coincides with the plane of the sky, i.e. if it is observed edge-on (Rybicki, 1987; Gerhard & Binney, 1996). This is a consequence of the ‘Fourier slice theorem’ or ‘projection-slice theorem’, which states that taking the Fourier transformation of the projection of a function ν along the line of sight is the same as taking the Fourier transform of ν first and then taking a slice of the Fourier transform along the line-of-sight. In the axisymmetric case, this leads to a ‘cone of ignorance’ within Fourier space that has an opening angle $90^\circ - i$. So-called ‘konus densities’ are densities whose Fourier transform are only positive within this cone. Therefore, they vanish in projection. The further away a galaxy’s viewing angle

is away from $i = 90^\circ$ the larger the cone of ignorance becomes and the larger the solution space for ν becomes.

This thesis adopts the *non-parametric*, axisymmetric implementation of Magorrian (1999) to explore plausible ν compatible with a given inclination (see Sec. 3). For ETGs the inclination is usually an unknown, therefore one must probe several viewing angles and compare the resulting dynamical models (Sec. 1.5). For a given trial inclination, the deprojection algorithm fits the observed surface brightness on an elliptical grid. The algorithm also includes several penalty terms that can be used to guide the solutions towards specific shapes (e.g. a boxy or discy shape) and avoid unrealistic, unsmooth solutions.

1.3.2 Spectroscopy

The second input required for the dynamical modelling are the (spatially resolved) stellar motions along the line-of-sight, known as the line-of-sight velocity distributions (LOSVD). The LOSVDs can be recovered from spatially resolved spectroscopic data like long-slit data or data from integral field unit (IFU) spectrographs. The latter is preferable because it provides the full 2D kinematic information on the sky, thus, increasing the constraints placed on the dynamical models. IFU data are stored in a 3D data cube (2D sky positions and 1D wavelength). The spatial pixels are binned further using a Voronoi tessellation method (Cappellari & Copin, 2003) with the goal to achieve a homogeneous and high Signal-to-Noise for each bin.

The spectra contain information about the LOSVDs because they consist of the integrated light/spectra of all stars in the galaxy. Each star has its own LOS velocity v and spectrum with absorption and emission features which are shifted by the (non-relativistic) Doppler effect following:

$$\lambda = \lambda_0 \cdot (1 + v/c) \quad (1.3)$$

where λ_0 is the rest-frame wavelength and λ the observed one. The superposition of all shifted stellar spectra then gives the observed integrated galaxy spectrum. In general, galaxy spectra can contain non-stellar features caused by interstellar gas or the accretion disc of supermassive black holes. For the dEs analysed in this thesis, however, the contribution from non-stellar features is negligible, and the spectra consist predominately by stellar absorption lines only. An example of such an integrated galaxy spectrum is shown in Fig. 1.11 with some of the most important absorption features marked.

The different velocities of the stars broaden and distort the shape of the absorption features in the integrated spectra. The resulting spectrum can be described by the *convolution* of the LOSVD with a ‘template’ stellar spectrum that is typical for the stellar population of the galaxy. To describe the template spectra of a galaxy, it is common to use stellar libraries (a collection of stellar spectra) such as ELODIE (Prugniel et al., 2007) or MILES (Falc3n-Barroso et al., 2011).

To recover the LOSVDs a ‘pixel-fitting method’ can be used which involves fitting the data pixel by pixel with a model spectrum generated by the convolution of a superposition of weighted template spectra with a model LOSVD. Implementations of this technique are

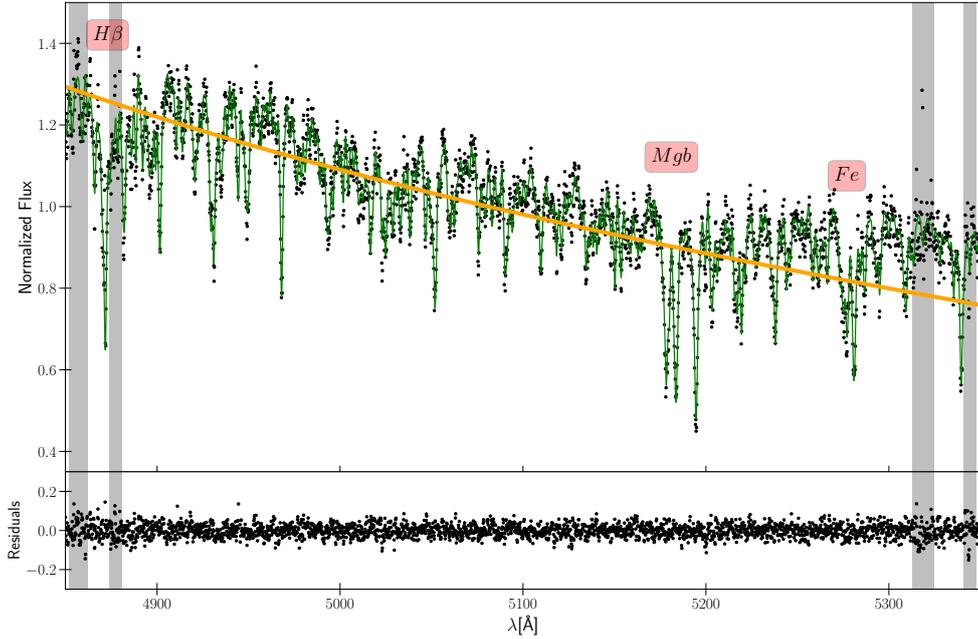


Figure 1.11: *Top panel:* Example of a kinematic fit to a Voronoi binned spectrum. In this case for a central bin of VCC 1861 (see Chapter 3). *Black dots:* The observed data. *Green:* The model spectrum. *Orange:* The continuum part of the model. *Bottom panel:* The corresponding residuals. The *grey* regions indicate regions that are masked during the fit. The location of the most important absorption lines is indicated.

pPXF (Cappellari, 2023) which allows the recovery of parametric LOSVDs (see below), BAYES-LOSVD (Falcón-Barroso & Martig, 2021), and the MPL-technique of Gebhardt et al. (2001) which characterize LOSVDs non-parametrically. The algorithm used in this thesis is WINGFIT (Thomas et al. in prep.) which allows both parametric and non-parametric descriptions of the LOSVDs.

Since both template spectra and model LOSVD are intertwined, it is crucial to provide an accurate description of the galaxy’s typical stellar spectrum. If the description is not accurate, this is called ‘template mismatch’. A detailed discussion of templates mismatch effects and how to avoid them can be found in Mehrgan et al. (2023).

It is common to describe parametric LOSVDs using a Gauss-Hermite expansion (van der Marel & Franx, 1993) up to an order n_{gh} :

$$\text{LOSVD}(x) = \gamma_0 \cdot e^{-\frac{x^2}{2}} \cdot \left(1 + \sum_{k=3}^{n_{gh}} h_k \mathcal{H}_k(x) \right) \quad (1.4)$$

with the relative, normalized velocity x :

$$x = \frac{(v - v_{\text{mean}})}{\sigma}, \quad (1.5)$$

and the Hermite polynomials defined as:

$$\mathcal{H}_k(x) = (-1)^k \cdot e^{x^2} \frac{d^k}{dx^k} e^{-x^2}. \quad (1.6)$$

Such a Gauss-Hermite description has $n_{gh} + 1$ free parameters: a normalization constant γ_0 , the mean velocity v_{mean} , the velocity dispersion σ , and the Hermite parameters h_3, h_4, h_5 etc. Using only the first 3 parameters ($\gamma, v_{\text{mean}}, \sigma$) eq. 1.4 becomes the standard Gaussian distribution which is enough to describe a simple net shift and broadening of the absorption/emission lines in the spectra. However, higher-order information is present in the absorption line shapes, which can be described using the Hermite coefficients h_k . The *odd* coefficients h_3, h_5, \dots characterize the *skewness* of the LOSVD while the *even* coefficients h_4, h_6, \dots its *kurtosis*. This is illustrated in Fig. 1.12. A positive h_3 means more stars are on prograde orbits (retrograde for negative h_3) which makes the LOSVDs asymmetric. In ‘ordinary’ ETGs a non-zero h_3 signal is often associated with an anti-correlated net rotation signal v_{mean} (Bender et al., 1994). A non-negative h_4 on the other hand signals symmetric deviations from a Gaussian: For a positive h_4 the LOSVD is sharply peaked in the centre and has significant high-velocity wings. For negative h_4 the LOSVD is more box-like with suppressed wings.

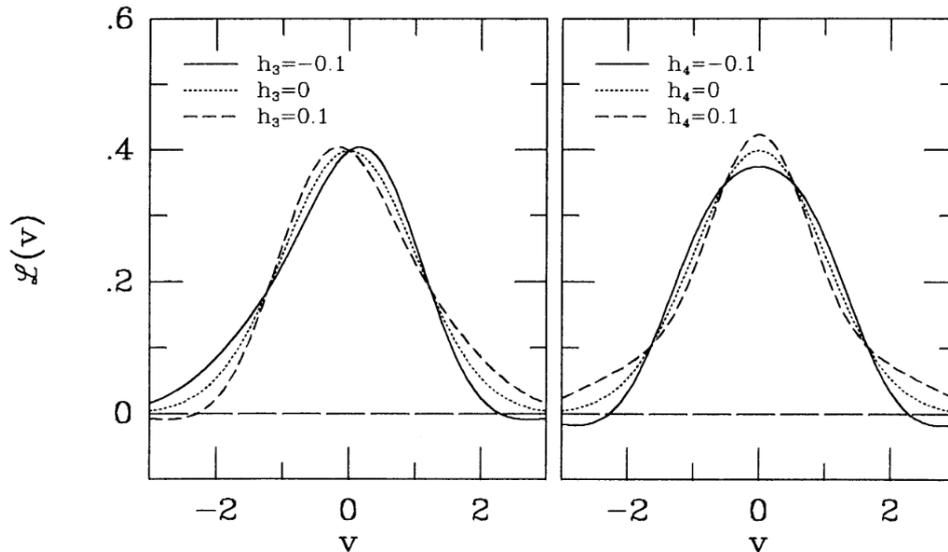


Figure 1.12: Illustration of the effect of a non-zero, odd Gauss-Hermite coefficient (*left panel*) and a non-zero, even Gauss-Hermite coefficient (*right panel*). Image credits: van der Marel & Franx (1993).

For ETGs the net rotation velocities³ are usually lower than their dispersions, which can reach up to $\lesssim 400$ km/s for the most massive ETGs and are $\sim 25 - 50$ km/s for the dEs that are the subject of this thesis. Therefore, one also speaks of dynamically *hot* systems.

³This means the velocity is measured within the rest frame of the ETG itself.

Higher order moments like h_3 and h_4 are typically close to zero, rarely exceeding ± 0.15 . Nonetheless, an accurate quantification of these deviations from a symmetric Gaussian distribution is essential if one wants to infer the 3D mass structure of galaxies because the mass and orbit structure are formally degenerate otherwise (see Sec. 1.4.1).

The prevalent approach in the literature is to parameterize LOSVDs up to the 4-th order h_4 . But more recent studies started to analyse higher-order moments as well (e.g. Liepold et al., 2020; den Brok et al., 2021; Thater et al., 2022; Quenneville et al., 2022) with the goal to exploit the full information contained in the spectra, similar to what can be achieved with non-parametric LOSVD model descriptions. However, it is not necessarily the case that a higher-order is also always the best option, since they also make the LOSVD models more prone to fitting noise that is prevalent in the spectra (see Sec. 1.5).

The same problem can happen for non-parametric LOSVD descriptions if the models are not sufficiently *smoothed*. Smoothing in non-parametric models is implemented by penalizing the goodness-of-fit χ^2 with some additional penalty term \mathcal{P} . The spectral fitting algorithm used in this thesis employs a second derivative smoothing term:

$$\mathcal{P} = \sum (\text{LOSVD}_{j-1} - 2\text{LOSVD}_j + \text{LOSVD}_{j+1})^2 \quad (1.7)$$

where the sum goes over all (equidistant) velocity bins j . This implementation correlates each LOSVD point with its two neighbouring points, compelling the curvature between them to be minimal. The strength of the penalty term and hence the amount of smoothing is controlled by a smoothing factor α_{smooth} . A high α_{smooth} means the LOSVDs is smoothed a lot, while $\alpha_{\text{smooth}} = 0$ implies no smoothing is applied. Like it is the case with the Gauss-Hermite parametrizations (with its maximum order n_{gh}) it is not obvious which value of α_{smooth} is best. Often these factors are fixed arbitrarily, other times Monte-Carlo simulations are employed to calibrate and find an appropriate value. Both these approaches are less than optimal, and it would be better to optimize the value of α_{smooth} from the data set it is actually being applied to.

Such a ‘data-driven’ method was established in Thomas & Lipka (2022), which is reprinted in App. A. The simple but versatile method allows one to optimize the penalty strengths of statistical models in general. It does so by employing the model selection tools that were first developed in Chapter 2 within the context of the dynamical modelling. Mathematically, this ‘data-driven’ method is a generalization of the famous Akaike model selection (see Sec. 1.5) and it is useful at several points in the modelling pipeline employed in this thesis. For the LOSVD recovery presented here, it allows one to derive the optimal values for n_{gh} and α_{smooth} directly from the observed spectra. This allows an optimal recovery for both Gauss-Hermite models and non-parametric LOSVD models without fitting the noise, all the while still extracting the entire information contained in the LOSVDs. A demonstration of this LOSVD recovery can be found in Thomas & Lipka (2022). Its application to the dynamical modelling itself is shown in Chapter 2.

The specific spectra of the dE sample analysed in this thesis were obtained with the VIRUS-W IFU spectrograph (Fabricius et al., 2008, 2012) mounted to the Harlan J. Smith telescope at the McDonald Observatory. The instrument achieves unprecedented spectral

resolution for dE galaxies. Thanks to this, the results presented in this thesis do not suffer from biased velocity dispersions like previous dE studies. The higher kinematic moments that this high resolution also allows to recover (for the first time for dEs) is essential information for the dynamical models as demonstrated in the next section.

1.4 Dynamical Modelling

The observed stellar motions trace a galaxy's gravitational potential $\Phi(\vec{x}, t)$ and, as such, its 3D total mass distribution $\rho_{\text{tot}}(\vec{x}, t)$. Given some assumptions (see below) one can therefore employ dynamical modelling techniques to infer the, a priori unknown, mass structure of various galaxy types.

Analogous to the treatment of gas in thermodynamics the large number of stars within a galaxy are unfeasible to model on an individual level, but the stellar system as a whole can be described well by its 6D phase-space distribution function (DF) $f(\vec{x}, \vec{v}, t)$, such that $f(\vec{x}, \vec{v}, t)d^3\vec{x}d^3\vec{v}$ is the probability that a star at time t has the velocity \vec{v} in the position \vec{x} .

Unlike small star clusters, galaxies can be treated as ‘collisionless’ systems, since their two-body relaxation times are much larger than the age of the Universe (Binney & Tremaine, 2008). This implies that the stellar tracers are not severely affected by individual two-body encounters with other stars, but instead are moving through the collective galaxy potential as a whole. Keeping generality, one can therefore describe the distribution function f using the collisionless⁴ Boltzmann equation (CBE):

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \nabla_x f \cdot \vec{v} - \nabla_v f \cdot \nabla_x \Phi \stackrel{!}{=} 0 \quad (1.8)$$

where ∇ is the nabla operator.

The total mass density distribution ρ_{tot} is connected directly to the gravitational potential by the Poisson equation:

$$\nabla_x^2 \Phi(\vec{x}, t) = 4\pi G \rho_{\text{tot}}(\vec{x}, t) \quad (1.9)$$

where G is the gravitational constant. In ‘self-consistent’ systems, which are systems where the potential Φ is generated solely by the particles that the DF f is describing (here stars), the DF f can be inferred directly if the phase-space coordinates of every star are known and the mass density ρ_{tot} can be inferred by integrating the DF over all velocities.

However, real galaxies are rarely treatable as ‘self-consistent’ systems: the observable DF describes only the visible components (e.g. stars or gas) but the total density ρ_{tot} is generated by both visible and dark components (e.g. dark matter or black holes). Nonetheless eq. 1.8 and eq. 1.9 hold, since the DF of the visible components is still tracing the total potential generated by the combined mass density. Consequently, the stars and their corresponding phase-space distribution can be used to constrain the potential.

⁴Similar to the collision terms, external force and diffusion terms are zero for galaxies.

On an abstract level, dynamical modelling essentially is the construction of DF models that emulate the observables of a galaxy, with the goal to infer its intrinsic properties, e.g. its 3D mass density ρ_{tot} . The most important observable in this context is the LOSVD of the stars, which is fully determined by its DF:

$$\text{LOSVD} = \int f(\vec{x}', \vec{v}') dv_{x'} dv_{y'} dz' \quad (1.10)$$

where \vec{x}' and \vec{v}' are now the *projected* coordinates such that the line of sight is parallel to the z' axis. Knowing the observed LOSVD one can then construct a number of model DFs, calculate and compare their model LOSVD to the observed one, and judge which model DF (and therefore density ρ_{tot}) is the most likely representation of the observed galaxy.

Over the past decades several dynamical modelling approaches have been employed to construct and represent the DFs of galaxies, starting from parametric DFs motivated by symmetry assumptions and physical reasoning (e.g. King, 1966; Wilson, 1975; Toomre, 1982; Hernquist, 1990; Dehnen & Gerhard, 1993; Gieles & Zocchi, 2015) towards more general approaches that will be discussed in the next two sections.

1.4.1 Jeans equations and the mass-anisotropy degeneracy

One can circumvent stipulating an explicit DF by approximating the CBE by taking its n -th order velocity order moments. This results in the set of ‘Jeans equations’ (Jeans, 1922; Binney & Tremaine, 2008).

Given some coordinate system (x_1, x_2, x_3) the *zero*-th order takes the form of a continuity equation:

$$\frac{\partial \varrho}{\partial t} + \sum_{i=1}^3 \frac{\partial(\varrho \langle v_i \rangle)}{\partial x_i} = 0 \quad (1.11)$$

where ϱ is the probability density of finding a star at a given position (i.e. the *normalized* mass density ρ_*) and $\langle v_i \rangle$ is the first order velocity moment at said position in the direction of x_i . These two properties are defined as

$$\varrho = \int_{-\infty}^{\infty} f d^3v, \quad (1.12)$$

and

$$\langle v_i \rangle = \frac{1}{\varrho} \int_{-\infty}^{\infty} v_i f d^3v. \quad (1.13)$$

Taking the *first*-order moment of the CBE results in:

$$\varrho \frac{\partial \langle v_k \rangle}{\partial t} + \sum_{i=1}^3 \varrho \langle v_i \rangle \frac{\partial \langle v_k \rangle}{\partial x_i} = -\varrho \frac{\partial \Phi}{\partial x_k} - \sum_{i=1}^3 \frac{\partial(\varrho \sigma_{ik}^2)}{\partial x_i} \quad (1.14)$$

for $k = 1, 2, 3$. This form of the Jeans equations takes advantage of the velocity dispersion tensor, which is defined as $\sigma_{ik}^2 = \langle v_i v_k \rangle - \langle v_i \rangle \langle v_k \rangle$.

While the line of sight components of the velocity moments are easily observable, generally this is not enough to *close* the Jeans equations⁵ since it is a set of 4 equations with 11 unknown quantities (ϕ , ϱ , $\langle \vec{v} \rangle$, $\hat{\sigma}^2$). Therefore, when constructing dynamical models for a galaxy (i.e. assuming some model form for ϱ and ϕ) one has to further restrict the solution space, e.g. by assuming a spherical symmetry or a well motivated form/subset of the velocity moments.

In a time-independent spherical system, for example, the Jeans equation simplifies (see Binney & Tremaine, 2008) to:

$$\frac{d(\varrho \langle v_r^2 \rangle)}{dr} + 2\frac{\beta}{r}\varrho \langle v_r^2 \rangle + \varrho \frac{d\Phi}{dr} = 0, \quad (1.15)$$

where $\langle v_r^2 \rangle$ are the *second* order velocity moments (compare eq 1.14), and the spherical anisotropy parameter β is defined by the velocity dispersions:

$$\beta(r) = 1 - \frac{\sigma_\theta^2(r) + \sigma_\phi^2(r)}{2\sigma_r(r)}. \quad (1.16)$$

Numerous modelling implementations/variations that make use of the Jeans equations have been used in the past to constrain the intrinsic structure of galaxies (e.g. Magorrian & Binney, 1994; Łokas, 2002; Cappellari, 2008; Loebman et al., 2012). The advantage of this approach is that it is fast and involves easily observable quantities. However, beyond the restricting assumptions about the velocity structure, solving the Jeans equations does not ensure physically possible DFs as f may be negative, nor does it employ the full observational information that may hide in the higher orders of the LOSVD.

Especially the latter is detrimental to an accurate and robust mass recovery. This is because the Jeans equation suffer from a ‘mass-anisotropy degeneracy’, which follows because the ‘line-of-sight’ dispersions σ_{LOS} (i.e. the observables) are affected by the intrinsic anisotropy distribution β and the 3D mass distribution ρ_{tot} in a degenerate manner (cf. Binney & Mamon, 1982; Bender et al., 1994; Gerhard, 1993; Read & Steger, 2017). The degeneracy implies that within the Jeans framework a change in the mass distribution of a galaxy can be entirely compensated by a corresponding redistribution of the typical orbit structure, i.e. by making the stars occupy more radial or tangential orbits on average.

To break this degeneracy, two things are required. *Firstly* additional information beyond just the line-of-sight dispersion and rotation, and *secondly* a more sophisticated dynamical modelling technique that is able to utilize this information.

Additional information can be proper motions of the stars (e.g. Strigari et al., 2007; Watkins et al., 2013), or the identification (and parallel modelling) of multiple different potential tracers (e.g. Walker & Peñarrubia, 2011; Napolitano et al., 2014). For the ETGs where such information is usually not available, the higher-order LOSVD information (cf. Sec. 1.3.2) provides a possible avenue to breaking the mass-anisotropy degeneracy (Gerhard, 1993; van der Marel & Franx, 1993). This is illustrated in Fig. 1.13 which shows

⁵Extending the equation system by taking higher velocity moments of the CBE does not solve this problem (cf. Binney & Tremaine, 2008).

the model LOSVDs for different anisotropy structures in identical potentials. The more radial/tangential the orbit structure becomes, the more peaked/flat-topped the LOSVD becomes. Measuring the kurtosis of the distribution (i.e. h_4) allows one to distinguish between models with different anisotropy structures, which narrows down the range of compatible anisotropy-mass pairs⁶ further than a gaussian descriptions allows. Combined with the spatial information (e.g. from IFU spectrographs) this allows meaningful constraints on the mass structure of galaxies.

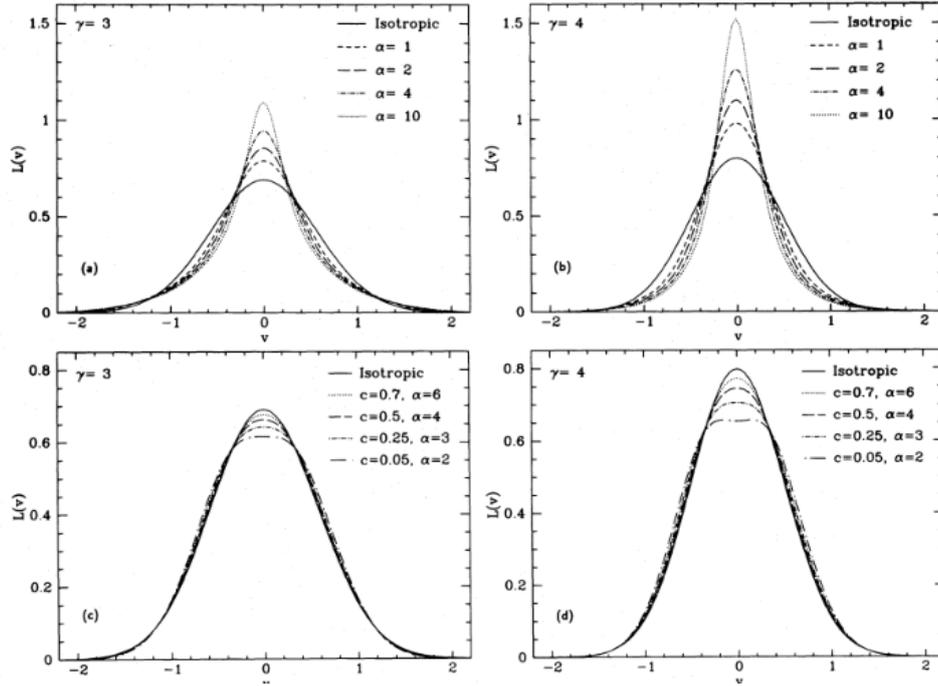


Figure 1.13: Illustration of the effects different anisotropy structure have on the higher moments of LOSVDs from Gerhard (1993). *Left panels:* For potentials with an outer stellar density distributions with slope $\gamma = -3$. *Right panels:* For the slope $\gamma = -4$. *Top:* Radially anisotropic orbit distributions. *Bottom:* Tangentially anisotropic distributions. For details of the parametrization of these models (density and anisotropy) see Gerhard (1993).

The second requirement to break the degeneracy is the use of dynamical modelling techniques that do not work in the restrictive framework of velocity dispersions (like the Jeans equations) but are defined more generally and are able to emulate the higher-order information. A very general dynamical modelling technique is to create n-body simulations that are tailored to the specific galaxy being investigated. In these made-to-measure (M2M) models (Syer & Tremaine, 1996; de Lorenzi et al., 2007; Long & Mao, 2010) the particles are given weights that are adjusted until the observations are matched by the corresponding

⁶For general systems there still exists no proof or guarantee that there is a *unique* mapping of LOSVDs to a single, specific mass-anisotropy pair.

n-body. The advantage of this technique is that it does not rely on restricting symmetry assumptions and that it is highly adaptive to the observables that are available. However, results can depend on the simulation setup (e.g. particle number, initial conditions) and the modelling is computationally very expensive since the particles of each n-body are evolved over several orbital periods. It is also not obvious when an M2M model is a good representation of the galaxy or when it is overfitting the noise in the data.

The second state-of-the-art modelling technique that retains a similar level of generality as the M2M technique but does not suffer from the same downsides is *orbit-superposition* modelling, often also called Schwarzschild modelling (Schwarzschild, 1979). The orbit-superposition modelling is able to use the full information contained within the LOSVDs and break the mass-anisotropy degeneracy without imposing too many restrictive assumptions. Therefore, it is the technique employed in this thesis to infer the majority of results for the dEs that are being investigated in Chapter 3 and 4. In the following, the basics of this approach are discussed.

1.4.2 Schwarzschild models

Instead of adjusting particle weights, like in the M2M-technique, the DF of a Schwarzschild model is modelled by a *superposition* of a set of orbits such that:

$$f = \sum w_i f_i \quad (1.17)$$

where f_i is the distribution function of a single orbit i .

To this end the technique takes advantage of the Jeans Theorem (Jeans, 1915; Binney & Tremaine, 2008) which states that, in steady-state systems, the phase-space density along stellar orbits is constant and that any DF that solves the CBE is a function of the *integrals of motions* I_n alone. Integrals of motions are functions of the phase-space coordinates that are conserved for all orbits. The most famous examples of such integrals are energy E and angular momentum L , which are integrals of motion in time-invariant and rotation-invariant systems.

From the Jeans Theorem it follows that in principle any steady-state solution of the CBE can be approximated by eq.1.17 as long as the space of integrals of motions is sampled sufficiently dense. This is accomplished by integrating a set of orbits (also known as an orbit library) in a given trial potential, where the initial conditions of the orbits are carefully chosen such that all plausible values for the integrals of motions are densely sampled.

This thesis makes extensive use of an advanced *axisymmetric* implementation of the Schwarzschild technique, which is based on the code of Thomas et al. (2004). In axisymmetric potentials $\Phi(r, \theta, \psi)$ orbits follow three integrals of motion: Energy E , the angular momentum in the z-direction L_z (if the z-axis is aligned with the symmetry axis), and a third, non-analytic integral of motion I_3 (Contopoulos, 1963).

The axisymmetry is exploited efficiently by computing the orbits in the 2-dimensional meridional plane (r, θ) and extending the model in the ϕ direction a posteriori. Each orbit model is then split and evaluated in spatial cells within said meridional plane. When

integrating the orbits in a given trial potential, the fraction of time an orbit has spent in a specific cell determines its contribution to the model's properties like luminosity and velocity moments in said bin (cf. Thomas et al., 2004).

Sampling E and L_z is achieved by tying it to the spatial library grid of the library. For each set of radii r_i and r_o (with $r_i \leq r_o$) of the spatial grid, one (or more) *equatorial* orbits is generated that has as its pericentre in the inner radial bin and its apocentre in the outer one (see also Richstone & Tremaine, 1988; Gebhardt et al., 2003; Siopis et al., 2009). The combination of each r_i and r_o provides a dense probing of the plausible energy and z-angular momentum space.

The third integral I_3 is then sampled for each pair of E and L_z by initialising the orbit integration from the zero-velocity-curve (ZVC) of the above grid points. The ZVC is defined by $v_r = v_\theta = 0$ and confines the phase-space region that is accessible to any orbit with a given energy E and L_z (Binney & Tremaine, 2008).

This can be seen in the radial 'surface of section' (SOS) that is spanned by r and v_r for example. Every time an orbit crosses the equatorial plane from 'below' (i.e. at $z = 0$, and $v_z > 0$) it leaves a mark in the surface of section that must be contained within the ZVC corresponding to the orbit's E and L_z . Orbits following the third integral I_3 do not randomly fill out the region bound by the ZVC, but are distributed along defined invariant curves. An example of such an invariant curve in the SOS is shown in Fig. 1.14.

From this, it becomes evident that in order to sample the third integral sufficiently dense for a given pair of E , L_z one should aim to fill out the corresponding ZVC. This can be achieved by *iteratively* filling gaps in a given SOS by adjusting the initial orbit conditions of subsequent orbits (for details Thomas et al., 2004) until the entire available phase-space is covered sufficiently.

For typical observations a total of several tens of thousands of orbits are generated to form a discrete basis of individual orbit DFs f_i from which principally any model DF that is compatible with the trial potential $\Phi(r, \theta, \psi)$ can be approximated by adjusting the orbital weights w_i .

In the context of dynamical modelling, one can use this DF-construction technique to constrain the intrinsic structure of observed galaxies by following 4 general steps:

- i) Establish a number of candidate mass models $\rho_{\text{tot}}(r, \theta)$ and calculate their corresponding gravitational potential $\Phi(r, \theta)$. Which and how many candidate models are being probed for a given galaxy is the modeller's choice and should be treated carefully (Lipka & Thomas in prep.). For ETGs for which gas and dust are negligible, the most popular choice is a 3 component model of the mass density:

$$\rho_{\text{tot}}(r, \theta) = \Upsilon \cdot \nu(r, \theta) + \rho_{\text{DM}}(r, \theta) + M_\bullet \cdot \delta(r) \quad (1.18)$$

where ν is the 3D stellar luminosity distribution and Υ its stellar mass-to-light ratio, ρ_{DM} is the (usually parametric) description of the dark matter halo density and M_\bullet is the mass of the central supermassive black hole (SMBH). In this 3 component model the luminosity distribution serves as boundary constrain and is obtained by *deprojecting* the observed photometry (see Sec. 1.3.1).

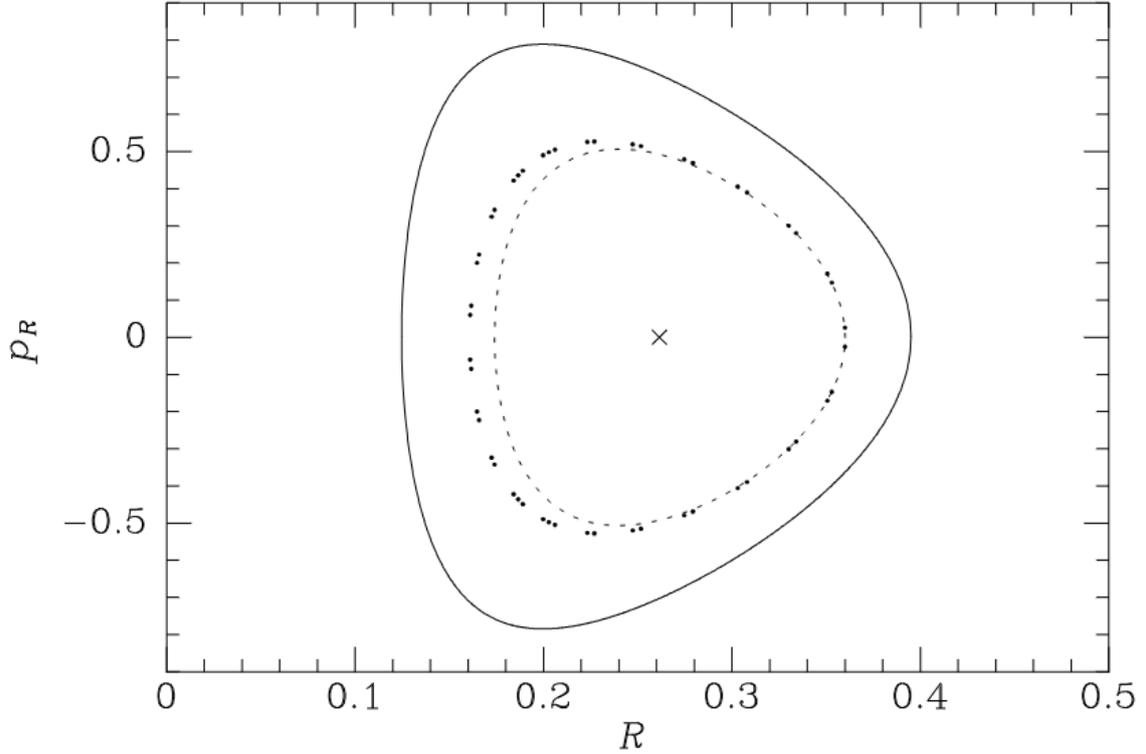


Figure 1.14: Example of a SoS in an axisymmetric potential from Binney & Tremaine (2008) who use the momentum notation p_r instead of velocity notation v_r . The dots show the equatorial passings of an orbit, while the solid line marks the corresponding ZVC in which any orbit with the same energy and z-axis angular momentum would fall.

- ii) For each trial potential, a DF *basis* f_i is constructed by integrating the orbits as described above and storing their properties in an ‘orbit library’ which contains the information of how much each orbit contributes to each spatial cell. Together with the freely adjustable orbit weights w_i the library determines all the model properties. Most important of which, its LOSVD, which can be conveniently written in matrix notation as:

$$\vec{L} = \vec{w} \cdot \mathbf{L}_{\text{orb}} \quad (1.19)$$

where the vector \vec{w} constrains the orbital weights w_i and the matrix \mathbf{L}_{orb} the contribution of the i -th orbit to the j -th LOSVD bin. In this notation the index j goes over all spatial and velocity bins of the observed data. Step ii) requires the definition of several nuisance parameters that define an ‘orbit library’, most importantly the number of orbits and its orientation in 3D space. The latter is defined by a set of viewing angles which, in the axisymmetric case, is a single angle: the inclination $i \in [0^\circ, 90^\circ]$. These nuisance parameters have not been treated with sufficient care in the past (see Chapter 2) even though they are formally just as important as the

mass/density parameters.

- iii) For each orbit library the optimum set of orbital weights \vec{w} is obtained by fitting the above model LOSVD vector to the corresponding observed one. However, in most cases this is non-trivial because the dimension of \vec{L} (i.e. the number of data constraints) is much lower than the dimension of \vec{w} (i.e. the number of free parameters/orbits). Therefore, a simple χ^2 fit would result in an overfitting of the data and a very unsmooth distribution function. To avoid this issue a regularization term is required which smooths the model DF and avoids overfitting. Following Richstone & Tremaine (1988) the code used in this thesis achieves this with a maximum-entropy term S . Instead of a simple χ^2 fit to the observed LOSVD the code maximises the quantity $S - \alpha\chi^2$ where S is defined as:

$$S = - \sum w_i \ln \left(\frac{w_i}{V_i} \right) \quad (1.20)$$

where the V_i are the phase volumes of the i -th orbit (see Thomas et al., 2004). Choosing to use the phase volumes in the denominator ensures that all orbits are a priori equally probable. However, one may use other bias factors instead to (a priori) bias the solution towards a specific form (cf. Neureiter et al., 2021). The regularization parameter α regulates the strength of this bias term. For small values, the entropy term dominates and the solution will be so smooth that it is unlikely to fit the observation well. For larger α on the other hand, the solution converges towards the χ^2 solution and the DF will be unsmooth and overfit the data.

- iv) Steps ii) and iii) are repeated for every candidate mass model that was established in step i), i.e. one finds an optimum set of orbit weights for every candidate model. Since the goal of dynamical modelling is to infer the intrinsic mass structure, one aims to identify the candidate model that best represents the galaxy that is being investigated. In the past this was done by comparing the different mass models based on their χ^2 , a by-product of the maximization of $S - \alpha\chi^2$. The mass model that achieved the smallest χ^2 out of all candidate models is then deemed to be the best *representation* of the observed galaxy. However, this χ^2 -approach neglects two fundamental issues. Firstly, α was often chosen rather arbitrarily and kept fixed for all candidate models, even though α is a regulator of over-/under-fitting and therefore strongly affects χ^2 . Secondly, different candidate models (step i) and nuisance parameter (step ii) result in different orbit trajectories. Consequently, every candidate model also has a *different* model basis (the elements of \mathbf{L}_{orb}) in the space of observables. On mathematical grounds, this means each candidate model establishes a separate statistical model with its own unique flexibility in the space of observables. And comparing the performance of *different* statistical models based alone on the goodness-of-fit χ^2 they achieved is not only wrong but biasing (Sec. 1.5). Fundamentally the Schwarzschild modelling technique is a model selection, therefore it is necessary to evaluate the performance of candidate models within a robust statistical

model selection framework. However, in the case of Schwarzschild models this is not trivial and new statistical techniques had to be developed (see Chapter 2).

In the course of this thesis, every step i)-iv) of the Schwarzschild technique, as well as the LOSVD recovery (Sec. 1.3.2), has been improved in preparation for the dE modelling. The improvements for each step are:

- i) A stellar mass-to-light ratio gradient $\Upsilon(r)$ was implemented (Chapter 3) which is a novelty for orbit-superposition models and has recently been introduced by (Mehrgan et al., 2024) for the modelling of giant ETGs. Allowing for a gradient in the stellar component is essential because the typical stellar populations can be expected to vary throughout a galaxy due to spatial changes in the star formation conditions (e.g. Parikh et al., 2018). In such a case, keeping Υ constant not only biases ρ_* but it also has reverberating effects on the dark matter and black hole recovery biasing those as well (via eq. 1.18). Similarly, a new model of the dark matter density ρ_{DM} has been implemented (Chapter 4) that is far less restrictive than the models that have been applied in the past. A non-restrictive Halo parametrization is essential if one wants to determine the dark matter profiles and shapes unbiasedly (Lipka & Thomas in prep.).
- ii) The effects and choice of nuisance parameters have been investigated carefully by applying various parameter setups to simulations (see Chapter 4) and by probing different values for the nuisance parameters in the dE modelling. Most importantly, this includes probing different inclinations i . In the past the vast majority of axisymmetric models has been modelled only in its edge-on ($i = 90^\circ$) projection because when probing different inclinations the edge-on projection generally resulted in the best χ^2 . In Chapter 2 it is demonstrated that this is due to a strong bias that stems from the naive application of χ^2 to select the best candidate model (step iv). It is shown how this inclination bias is not present if the orbit models are treated within the model selection framework developed here in this thesis. Not only does this enable dynamical inclination recovery, but it also improves the mass and anisotropy recovery in general because the model is now projected to the space of the observables using the correct viewing angle.
- iii) Analogous to the penalty strength α_{smooth} in the LOSVD recovery, it is crucial to find the optimal regularization strength of the orbit models α , such that both under- and over-fitting are avoided. In past applications of the Schwarzschild technique, the choice of this penalty strength was often made ‘by eye’ or, at best, using Monte-Carlo simulations. Like it was the case for the LOSVDs smoothing, the ‘data-driven’ method that was developed within this thesis (Chapter 2) and Thomas & Lipka (2022) allows the optimization of α directly from the data. This not only vastly improves the constraining power of the Schwarzschild modelling technique but allows for an unbiased recovery of the mass parameters (Chapter 2).

- iv) The model selection framework is not only useful for the correction of the inclination bias and the smoothing optimization of α_{smooth} and α . In Chapter 2 it is demonstrated that the selection of the best candidate model on a fundamental, statistical level is a model selection problem and not a naive parameter estimation problem as thought previously. When comparing the performance of different mass models established in step i) one must take into account that their intrinsic flexibility is not identical. A new approach was developed in this thesis that allows the quantification of this intrinsic flexibility. This, for the first time, allows an evaluation of the Schwarzschild models within a model selection paradigm. In statistics, various model selection criteria have been proposed. For the application to Schwarzschild models and the smoothing optimization, a new criterion had to be developed. The new AIC_p criterion is a generalization of the Akaike criterion (AIC). The basics of the classical AIC are presented in Sec. 1.5, the generalized criterion is the subject of Chapter 2 and Thomas & Lipka (2022).

1.5 The Basics of Model Selection and Information Theory

Scientific inquiry, like it is described in the last section, necessitates extracting hidden information (e.g. the 3D density ρ_{tot}) that is not directly observable by means of *fitting* a model of the underlying process to the available observables. However, in many applications, it is a priori unclear which model is an adequate description of the physical process that generated the data. In this case, several trial/candidate models can be established and fitted to the same data. Then their performance can be evaluated and one can decide which model is most suitable in the description of the physical process.

A model selection technique is a statistical tool to choose the best model out of a pool of those trial models. A widely known philosophical example of this is ‘Occam’s razor’ which states that, generally, the simpler hypothesis (or model) that is able to explain a phenomenon should be preferred over a more complicated one. This is an intuitive principle. For example, when fitting the same set of data, once with a linear model and once with a parabolic model, the latter will *always* achieve a nominally better fit (i.e. a smaller χ^2), but this doesn’t mean the underlying process is not a linear one.

Generally speaking, more complex models (i.e. with more free parameters) will always be able to better adapt to any perturbation in the data. But this is only desirable if the perturbation is caused by a real signal, not by random noise. This is illustrated in Fig. 1.15, which shows the fits of 4 different statistical models to a mock LOSVD. The highly flexible models in the top row fit portions of the noise. Using these models one would *overfit* the data, i.e. detect artificial features in the LOSVD. Conversely, the models in the bottom row ‘underfit’ as they are not flexible enough to recover the more complex truth.

The approach to selecting the best model which neither over- nor under-fits is therefore:

- i) Determine the ability of each trial model to fit the data set, usually by calculating

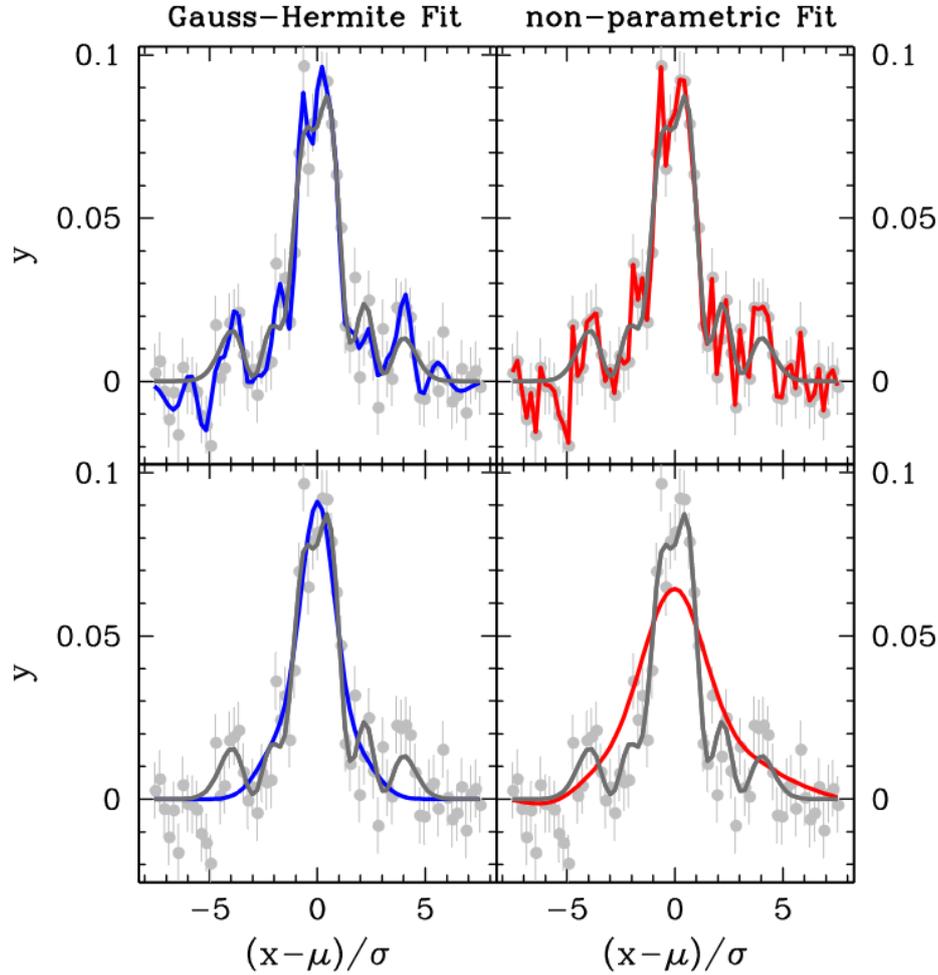


Figure 1.15: An exaggerated example of over- and underfitting for 4 statistical models from Thomas & Lipka (2022). The *grey* curves are the truth (i.e. the noise-free data-generating process). The grey dots are the *noisy* data. In *blue* are 2 parametric models (Gauss-Hermite Fits) and in *red* 2 non-parametric models. Models in the top panels overfit the data, while the models in the bottom underfit, meaning they are not able to recover the complexity of the truth.

the χ^2 of each model.

- ii) Determine the complexity/flexibility of each trial model, usually quantified in terms of its number of free fit parameters m .
- iii) Judge each model using a model selection *criterion* that balances the above (χ^2 and m) and finds the sweet-spot between simplicity and fit quality.

Step i) is straight-forward and, in the context of dynamical modelling, the currently accepted approach to ‘select’ the best dynamical model (cf. Sec. 1.4). This is a valid

approach only if all trial models have *identical* model flexibility, i.e. if step ii) can be neglected. Only then step iii) becomes redundant and a χ^2 -only selection is unbiased.

However, neither for Schwarzschild models, nor any other dynamical modelling technique, the assumption that all trial models have identical flexibility has actually ever been investigated or verified. This has two reasons. Firstly, because it was not obvious that Schwarzschild models should be treated in a model selection framework in the first place, even though the different trial models consist of different orbits and thus different basis vectors. Secondly, and more importantly, because there was no known approach to calculating, or at least estimating, the flexibility of Schwarzschild models. It is not as straightforward to simply equate the number of orbital weights w_i to that of the number of *free* parameters, because, strictly speaking, this only holds for *linear* statistical models without any a priori restrictions on the parameters (Burnham & Anderson, 2002; Andrae et al., 2010). Schwarzschild models, however, are neither linear nor are their parameters unrestricted. Instead, they are regularized models with a penalty term (eq. 1.20) and their parameters w_i are demanded to be non-negative (otherwise the DF could be negative).

For these reasons it continued to be unclear whether the assumption that all trial models have the same flexibility is correct, and hints that different trial Schwarzschild models have vastly different flexibility were noticed but remained unresolved. One such hint was that a suspicious amount of times, *edge-on* orbit models provided a much better χ^2 than inclined models (e.g. Thomas et al., 2007; Onken et al., 2007). This could be explained if edge-on models have higher flexibility than inclined models.

If one can establish an alternative approach to measuring the flexibility of the dynamical models (as is done in this thesis) and find that the flexibility of the orbit models varies, the obvious conclusion is that step iii) is imperative to select the best orbit/mass model. But this raises a second question: which specific model selection criterion is most suitable in the context of dynamical modelling?

A manifold of different model selection criteria exist in statistics (Burnham & Anderson, 2002). Among the most popular ones are the Akaike Information criterion (AIC) and the Bayesian information criterion (BIC) (Akaike, 1973; Akaike, 1974; Schwarz, 1978). When written in terms of goodness-of-fit⁷ χ^2 , the complexity m , and the number of data points N , they read as

$$\text{AIC} = \chi^2 + 2m, \quad (1.21)$$

and

$$\text{BIC} = \chi^2 + \ln(N) \cdot m. \quad (1.22)$$

Both criteria penalize the fit quality χ^2 by the complexity m of the model, but they differ in how much they weigh these two factors. BIC typically penalizes complexity more strongly (a factor of $\ln(N)$ vs 2). Again, note that the complexity m only equals the number count of fit parameters for linear models without prior restrictions. The two criteria hold for maximum-likelihood models and large N , but a generalized Akaike criterion that extends to penalized models can be derived (this is done in Thomas & Lipka, 2022).

⁷Note that using χ^2 instead of the log-likelihood assumes errors follow a normal distribution.

While BIC is (typically) derived within a Bayesian framework, AIC is an *information* criterion in the stricter sense of the word because it is derived from an information theoretical approach. Fundamentally, AIC is an attempt to minimize the information loss when modelling the ‘truth’ f using a model g . In information theory the relative information loss between two probability functions is quantified using the *Kullback-Leibler-distance* defined as:

$$\text{KB}(f, g) = \int f(x) \cdot \ln \left(\frac{f(x)}{g(x|\theta)} \right) dx \geq 0 \quad (1.23)$$

where $g(x)$ is the model that one wants to approximate the truth f with. The smaller $\text{KB}(f, g)$ the closer is the model to truth. Therefore, the target of model selection should be to minimize the Kullback-Leibler-distance between f and g .

However, calculating the Kullback-Leibler distance between the truth and the model directly is impossible, because neither the truth f nor the true parameters θ_0 of g that yield the best approximation of f are known (one only estimates θ_0 from noisy data). Therefore, the ansatz of AIC is to minimize the *expected, estimated* information loss, which can be approximated as eq. 1.21 as Akaike showed.

1.6 Goals and outline of this thesis

Dynamical modelling is the key to unveiling the mass and dynamical structure of galaxies. This information, that goes well beyond a galaxies morphological appearance, provides strong constraints on galaxy evolutionary mechanisms and structure formation in the Universe. This in turn may allow probing predictions from cosmological theories and ascertain which processes are responsible for the large diversity in morphological appearance that are being seen in the galaxies populating the Hubble diagram.

In the context of dwarf ellipticals, we aim to understand whether the apparent connection with dwarf spirals (e.g. in terms of their surface brightness distribution) is an indicator of a deeper evolutionary connection and, if so, what has led some dwarf galaxies to end up as the quiescent dEs while others are still actively forming stars. While direct observables like the surface brightness distribution are well studied for dEs, little is known about their intrinsic structure. It is still open question whether a connection between dEs and dwarf spirals can also be found in their orbit structure (angular momentum and anisotropy), their dark matter halo, and black hole masses.

This thesis sets out to answer these questions for a sample of dEs located in the Virgo cluster. It is the first time that the structure of dEs is studied with sophisticated dynamical models that are able to employ the full kinematic information hidden in the spectra and do not suffer from intrinsic degeneracies and restricting assumptions. This required an unprecedented IFU data set with very high spectral resolution. Only this enabled the first measurement of the higher-order LOSVD moments for dEs, which is essential information for the construction of the dynamical models. Numerous improvements to the Schwarzschild modelling pipeline itself were developed throughout this thesis, ensuring a robust and unbiased recovery of the structure of dEs.

The goals of this thesis are two-fold. Firstly, improve the dynamical modelling inference generally, by placing it in a rigorous statistical model selection framework. And secondly, use these improvements in the modelling to address a set of science questions concerning the origin and evolution of dwarf elliptical galaxies. Therefore, the thesis is structured as follows:

- Chapter 2 introduces the new model selection framework and applies it to the dynamical modelling machinery. It lays out the arguments why Schwarzschild modelling *fundamentally* is a model selection problem, and not a parameter estimation problem. Therefore, varying model flexibility between different Schwarzschild models can bias galaxy constraints. The chapter presents a novel approach to estimate the flexibility of the dynamical models using bootstrap iterations, and demonstrates how their flexibility systematically changes with various model properties like inclination, mass-to-light ratios or angular momentum. This leads to biases in the χ^2 surfaces. In the case of the inclination it even hinders any meaningful constraints. The chapter shows how one can use the estimated flexibility to correct the χ^2 values, achieving an unbiased model evaluation. The new model evaluation is calibrated and tested on 10 different simulated galaxies. The chapter also shows the first application of the new procedure to a real galaxy (NGC 3368). The inclination of this disc galaxy is known to be $i \approx 53^\circ$ from external constraints. While the old χ^2 approach favours the model to be edge-on ($i = 90^\circ$) the new model evaluation correctly recovers the inclination. Finally, the chapter also presents how the regularization parameter α of the Schwarzschild models can be optimized directly from the data using the new model selection framework. This drastically improves the accuracy of the mass and orbit structure recovery. The chapter was published in Lipka & Thomas (2021). A thematically closely related article that explores some findings of Chapter 2 in a more general, statistical framework and underpins its mathematical foundations is Thomas & Lipka (2022).
- Chapter 3 and Chapter 4 employ the new and improved modelling machinery to tackle various science questions concerning dwarf galaxies. A total of 9 dEs associated with the Virgo cluster is modelled and analysed comprehensively. In the first part of Chapter 3 the sample data is presented, and the modelling pipeline is discussed in detail. In the second part of Chapter 3 the *stellar* structure (3D mass and orbit structure) is presented. In a comparison with the published literature, the chapter also demonstrates that high resolution spectra are indispensable if the low velocity dispersion are to be recovered unbiasedly. The study represents one of the first IFU studies of dEs that analyses higher LOSVD moments, and dynamically models dwarf ellipticals. Supplementing the constraints on the intrinsic structure of the dEs provided by the dynamical models, a stellar population analysis is also performed, providing information about the stellar age, metallicity, and abundance ratios of the dEs. Combined with the dynamically recovered mass-to-light ratios (and gradients) the population results pose constraints on the initial mass function of the

stellar populations. Beyond the stellar mass distributions, their angular momentum and intrinsic anisotropy structure is studied. The results shown in chapter 3 were published in Lipka et al. (in press). Chapter 4 focuses on the dynamically recovered dark matter halos and black holes of the dE sample and was originally published in Lipka et al. (in press). In the chapter, it is described how one can optimize the construction of halo models and how the dynamical results should be interpreted such that potential biases from the modelling are avoided. Dark matter fractions, dark matter slopes and, for the first time, the flattening of the halos are recovered. It is investigated whether these observational constraints are in tension with predictions from standard Λ CDM cosmology. Finally, combining the results from chapter 3 and chapter 4, it is explored whether dEs are the remnants of transformed spirals and, if so, which transformation processes are the most likely culprits.

Chapter 2

A novel approach to optimize the regularization and evaluation of dynamical models using a model selection framework

Bibliographic information

Lipka Mathias, Jens Thomas - A novel approach to optimize the regularization and evaluation of dynamical models using a model selection framework, Monthly Notices of the Royal Astronomical Society, Volume 504, Issue 3, pp.4599-4625
DOI: 10.1093/mnras/stab1092

The Author's contribution

As first author, I wrote all sections and produced all figures in the manuscript myself. The final formulation of a generalised model selection framework and the application to optimise the dynamical models and regularization was developed in frequent discussions between myself and Jens Thomas. The initial idea to measure the flexibility of penalised models using bootstraps and to apply ideas of model selection to solve known biases in Schwarzschild Models was conceived by Jens Thomas. It is founded in Thomas & Lipka (2022), a closely related article that is appended to this thesis (App. A). In it, the generalized version of the Akaike information criterion is developed from a rigorous mathematical standpoint, and it is illustrated how the approach is useful in statistical modelling in general. First tests regarding the inclination recovery were made during my master thesis. Those were redone or improved after subsequent refinements of our techniques and implementation. In particular, the introduction of the generalized Akaike information criterion and the optimization of the regularization parameter improved all results significantly. I generated all simulations and mock observations and produced all dynamical models. The

Schwarzschild code with the implementation of the generalised model selection was provided by Jens Thomas. For the modelling of the galaxy NGC 3368 I used the pre-reduced data sets of Nowak et al. (2010) as input for the modelling. Jens Thomas provided invaluable support to this work in the form of discussions, ideas, and revisions of the manuscript. Roberto Saglia provided feedback for the final version of the manuscript.

Copyright

©2021 MNRAS
reprinted on pages 37- 63

A novel approach to optimize the regularization and evaluation of dynamical models using a model selection framework

Mathias Lipka^{1,2★} and Jens Thomas^{1,2}

¹Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching, Germany

²Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany

Accepted 2021 April 14. Received 2021 April 14; in original form 2021 February 9

ABSTRACT

Orbit superposition models are a non-parametric dynamical modelling technique to determine the mass of a galaxy's central supermassive black hole (SMBH), its stars, or its dark matter halo. One of the main problems is how to decide which model out of a large pool of trial models based on different assumed mass distributions represents the true structure of an observed galaxy best. We show that the traditional approach to judge models solely by their goodness-of-fit can lead to substantial biases in estimated galaxy properties caused by varying model flexibilities. We demonstrate how the flexibility of the models can be estimated using bootstrap iterations and present a model selection framework that removes these biases by taking the variable flexibility into account in the model evaluation. We extend the model selection approach to optimize the degree of regularization directly from the data. Altogether, this leads to a significant improvement of the constraining power of the modelling technique. We show with simulations that one can reconstruct the mass, anisotropy, and viewing angle of an axisymmetric galaxy with a few per cent accuracy from realistic observational data with fully resolved line-of-sight velocity distributions (LOSVDs). In a first application, we reproduce a photometric estimate of the inclination of the disc galaxy NGC 3368 to within 5° accuracy from kinematic data that cover only a few sphere-of-influence radii around the galaxy's SMBH. This demonstrates the constraining power that can be achieved with orbit models based on fully resolved LOSVDs and a model selection framework.

Key words: methods: statistical – galaxies: individual (NGC 3368) – galaxies: kinematics and dynamics – galaxies: structure.

1 INTRODUCTION

Revealing the internal structure of external galaxies is a challenging, yet essential task for a broader understanding of galaxy evolution. Since observations of external galaxies are limited to the projected kinematics and the surface brightness of the luminous galaxy components the common approach is to establish dynamical models of an observed galaxy and to compare their projections to the corresponding observations. A collisionless galaxy component, such as the ensemble of its stars, follows the collisionless Boltzmann equation and can be fully characterized by its phase-space distribution function. Therefore a dynamical model can, at least in principal, describe such a system of stars if it can sufficiently emulate the real distribution function. If a galaxy could not be assumed to be in equilibrium, dealing with all the freedom in the distribution function during the modelling process would be hopeless. However, in equilibrium the Jeans Theorem can be invoked (e.g. Binney & Tremaine 2008) and the distribution function has a relatively simple structure which can be expressed as a function of the integrals of motion meaning it takes the form of an orbit superposition.

Even though stellar systems with a high degree of symmetry can sometimes be modelled well using analytic distribution functions, a more universal approach to dynamical modelling is generally re-

quired to deal with the full range of compatible distribution functions. Schwarzschild modelling is an efficient numerical approach based on the superposition of stellar orbits to construct such dynamical models. In its initial version, established by Schwarzschild (1979), such orbit models were designed to replicate a given triaxial density distribution in a self-consistent manner. In the subsequent decades the Schwarzschild technique was extended to include the fitting of kinematic observations (e.g. Levison & Richstone 1985; Rix et al. 1997; Cretton et al. 1999; Thomas et al. 2004; van den Bosch et al. 2008), making it possible to constrain the distribution function of an observed galaxy more tightly and enabling the estimation of intrinsic properties of stellar systems such as black hole mass (e.g. van der Marel et al. 1998; Gebhardt et al. 2000, 2003; Cappellari et al. 2002; Rusli et al. 2013), mass-to-light ratios (e.g. Thomas et al. 2005; Cappellari et al. 2006; Thomas et al. 2011), dark matter haloes (e.g. Thomas et al. 2005, 2007b, 2009b; Rusli et al. 2013; Leung et al. 2018), the velocity dispersion anisotropy and orbital structure (e.g. van de Ven, de Zeeuw & van den Bosch 2008; Thomas et al. 2009a, 2014; Kowalczyk, Łokas & Valluri 2017), and the intrinsic shape of the stellar distribution (e.g. Jin et al. 2020).

Several different implementations of the Schwarzschild method with varying degrees of symmetry have been described (e.g. Rix et al. 1997; Cretton et al. 1999; Gebhardt et al. 2000, 2003; Häfner et al. 2000; Siopis & Kandrup 2000; Thomas et al. 2004; Valluri, Merritt & Emsellem 2004; van den Bosch et al. 2008; Vasiliev & Athanassoula 2015; Vasiliev & Valluri 2020; Neureiter et al. 2021),

* E-mail: mlipka@mpe.mpg.de

but the main steps of the method are rather general. Very briefly: If an assumed gravitational potential is given, one needs to compute tens of thousands of representative orbits in the available phase space. These orbits are combined to a galaxy model, where each orbit can carry an adjustable number of stars. These so-called orbital weights (or orbital occupation numbers) together with each orbit's intrinsic and projected properties then determine the dynamical model. Through optimization of the orbital weights, the model can be adapted to the observations. In this sense, the orbital weights are the variables of the model. While different algorithms are used for this weight optimization, they are all tied to a χ^2 -minimization framework.

With the rise of integral-field spectrographs in the last decade, the amount of kinematic data that can be used to constrain these models has increased significantly. While many early studies had to rely on line-of-sight velocity distributions (LOSVDs) that were parametrized by Gauss–Hermite series expansions (Gerhard 1993; van der Marel & Franx 1993) up to fourth order (often along a 1d slit; Bender, Saglia & Gerhard 1994), it is now often possible to reliably obtain higher Gauss–Hermite moments (e.g. Krajnović et al. 2009; Liepold et al. 2020) or measure the fully resolved, non-parametric LOSVDs over the 2d sky area occupied by the galaxy (Mehrgan et al. 2019). The number of measured data points per galaxy has thus increased from a few dozen to several thousands. Still, one of the characteristics of the Schwarzschild method is that one is usually faced with a situation where the number of free model parameters is significantly larger than the number of observational data points, mainly because the number of orbits is typically very large and each orbit is associated with a free orbital weight parameter. For simple, linear models this would imply a negative number of degrees of freedom of the resulting χ^2 distribution (Press et al. 1992) and the solution of the χ^2 -minimization is non-unique (cf. Valluri et al. 2004; Magorrian 2006; Neureiter et al. 2021). Under the assumption that a model is generally detailed enough to deal with all the structure in the data, one would consequently expect that every data set is perfectly fitted due to the models' comparatively huge number of free parameters, resulting in a $\chi^2 \rightarrow 0$. In practice, this does not happen for two reasons. First, data with a realistic amount of noise cannot be fitted perfectly when the orbital weights are constrained to be non-negative (see Magorrian 2006 for a discussion). Secondly, a perfect fit is suppressed when a regularization term is applied in the weight optimization, e.g. via the commonly used maximum penalized likelihood or maximum entropy approaches. The smoothing induced by such a regularization term allows to discard physically implausible solutions and prevents overfitting by effectively reducing the flexibility of the dynamical model. Despite the differences in the individual implementations of regularization, the result is always the same: The χ^2 of the regularized model is significantly larger than zero even though the model's parameters typically outnumber the data points. This suggests that the *effective* number of free parameters is smaller than the nominal number of variables in the model, implying a reduction in the model's flexibility (i.e. its ability to fit noisy data). Beyond the regularization another core aspect in reducing the flexibility of Schwarzschild models is the prior constraint on the orbital weights to be non-negative to avoid negative phase-space densities in the model.

The reduction of model flexibility is a generic property of regularized models, of models with prior constraints imposed on their free parameters, or in general of nonlinear models (cf. Andrae, Schulze-Hartung & Melchior 2010). As a consequence, because the *effective* number of parameters is an unknown, the interpretation of the absolute value of χ^2 becomes less obvious. This is particularly important, because in most Schwarzschild applications, the primary

interest is not in the distribution of the orbital weights. Instead, one aims to determine the mass composition of a galaxy, which requires the comparison of a number of Schwarzschild models that were obtained with different trial gravitational potentials. Traditionally this comparison is done by evaluating the χ^2 values inherited from the orbital weight optimization of each trial model. Given the fact that these χ^2 values depend on the model flexibility (or its *effective* number of free parameters) it may be important to take this into account in the evaluation process.

The goal of this paper is threefold. First, we provide a method to estimate the flexibility of Schwarzschild Models, or the *effective* number of free parameters respectively, that can robustly deal with all the smoothing constraints, positivity constraints and nonlinearities of the models. Secondly, we demonstrate that varying model flexibilities lead to biased χ^2 surfaces when evaluating orbit models obtained using different mass distributions, viewing angles or assumed regularization powers. Thirdly, we provide a model selection framework that allows to overcome these biases that can arise in the simple χ^2 comparison framework. The new framework not only improves the constraining power of Schwarzschild models significantly, but also enables a new data-driven approach to optimize the amount of regularization. All these considerations can probably be generalized to other nonlinear, non-parametric, non-dynamical models that involve regularization or prior constraints.

As the main application example, we investigate the reconstruction of the viewing angle (or equivalently the intrinsic flattening) under which an axisymmetric galaxy is observed. The question whether it is feasible to constrain the inclination/intrinsic shape via dynamical modelling is as of yet a point of contention. In early works Verolme et al. (2002) argue that they constrained the inclination of M32 to $70^\circ \pm 5^\circ$ using axisymmetric, three-integral, Schwarzschild models. However, follow-up results of Krajnović et al. (2005), who tried to recover the inclination of semi-analytic galaxy models with axisymmetric models, suggest that different inclinations are degenerate even under ideal conditions. Cappellari et al. (2006) confirmed these results for a large sample of early-type galaxies in so far as they state that Schwarzschild models with a wide range of inclinations are able to fit the observed kinematics well within the errors. Onken et al. (2007) modelled the bulge of the Seyfert 1 galaxy NGC 4151 edge-on and with an inclination $i = 23^\circ$ obtained from the observed ellipticity of the large-scale disc, with the result that the edge-on model provides better fits to the kinematic constraints than the 23° model, suggesting either that the bulge is not aligned with the disc or that the inclination recovery using dynamical modelling is biased. Similarly, Thomas et al. (2007b) found that a surprising amount of the 17 early-type galaxies they dynamically modelled with axisymmetric three-integral models are fitted best when the orbit model is viewed edge-on and they argue there may be a small inclination bias favouring edge-on Schwarzschild models. In this paper we will demonstrate that the assumed inclination strongly affects the flexibility of orbit models which could explain the observed edge-on preference of previous studies. We will show that the inclination of axisymmetric galaxies can be well constrained from typical observational data with the new model selection framework as it considers the model flexibility in the evaluation of the fitted models.

The paper is organized as follows: In Section 2 we outline our motivation for using a model selection approach in the evaluation of Schwarzschild models in detail. In Section 3 we lay out two bootstrap methods which facilitate the quantification of a model's intrinsic flexibility. A discussion of an evaluation approach based on model selection techniques follows in Section 4. We then test the model selection with respect to the inclination reconstruction

on a number of simulated galaxies in Section 5. In Section 6 we extend the analysis to mass parameters. Section 7 further refines our approach by including the regularization power. Equipped with all these insights we model the disc galaxy NGC 3368 in Section 8 with our refined model selection framework and set it into contrast to the prevalent χ^2 -minimization. Section 9 discusses general aspects related to modelling degeneracies and the inclination recovery. The paper concludes with a summary in Section 10.

2 MOTIVATING A MODEL SELECTION APPROACH TO THE SCHWARZSCHILD TECHNIQUE

As mentioned in the Introduction above, we want to demonstrate the power of the model selection framework using the example of Schwarzschild orbit superposition models. The main reasons for picking up Schwarzschild models are as follows: (i) This technique is very general and allows to model any kind of galaxy without apriori assumptions upon the orbital structure. (ii) The method is observationally not restricted to moments of the velocity distribution of the stars but instead can deal with the full information contained in the LOSVDs (e.g. Thomas et al. 2004; Mehrgan et al. 2019; Vasiliev & Valluri 2020; Neureiter et al. 2021). The specific implementation of the Schwarzschild method that we consider in this paper assumes axially symmetric galaxies and is described in detail in Thomas et al. (2004).

In Schwarzschild models, one first specifies a trial mass distribution and viewing angle via parameters like the stellar mass-to-light ratio Υ , black hole mass M_\bullet , dark-halo parameters and inclination i . Then one calculates an orbit library with thousands of orbits whose projected properties form the building blocks of the superposition model. By computing the best-fitting orbital weights one can assess how well the trial mass distribution allows to explain the observational data. A systematic search through the space of trial mass models then leads to the identification of a best-fit mass distribution for each galaxy. This is only a very brief sketch of the method (we provide a more detailed description in App. A and Section 2.1) but it already demonstrates one important aspect: the fact that one *first* needs to specify a set of mass parameters (and the inclination) *before* the orbit distribution can be determined will turn out to be significant. As a consequence it is helpful to conceptually distinguish two distinct parameter layers that together define the properties of every Schwarzschild model: Minimization parameters (orbital weights w_i) and selection parameters (all other parameters, i.e. M_\bullet , Υ , i , ...). It is the latter parameters that require an evaluation using a *model selection* framework.

We present our reasoning for this distinction of parameter layers in Sections 2.1 and 2.2. We also recapitulate our regularization approach to finding the optimum minimization parameters (i.e. orbital weights) in 2.2. The theoretically less interested reader may skip ahead to Section 2.3 where we show that a simple χ^2 evaluation where all parameters are treated equally can lead to a bias, using the inclination as an example. As we will show throughout this paper, other selection parameters such as the stellar mass can also suffer from such a bias.

2.1 Selection versus minimization parameters

It is tempting to consider the question of identifying the best-fit Schwarzschild model in terms of a huge $\chi^2(M_\bullet, \Upsilon, i, \dots, w_i)$ minimization, where all the parameters are treated equally as free parameters of a *single 'global' Schwarzschild model*. This concerns mass parameters like M_\bullet and Υ , viewing angles like the inclination

i and the orbital weights w_i (cf. App. A). It is therefore common practice to assume that the best choice for $M_\bullet, \Upsilon, i, \dots$ is given by the trial set that resulted in the smallest χ^2 without much concern of the role of orbital weights.

Using a sophisticated Bayesian framework Magorrian (2006) acknowledges that there often exist many possible combinations of orbital weights (or equivalently distribution functions) that are consistent with given observational data and an assumed trial potential. Magorrian (2006) argues that a more appropriate approach sums over all *possible* distribution functions for a given potential. This is achieved by marginalizing over the orbital weights and weighing the corresponding likelihoods by a suitable prior. Magorrian (2006) argues that this approach allows a more accurate calculation of the likelihood for a given potential, and that the odds of one set of trial parameters ($M_\bullet, \Upsilon, i, \dots$) over another can be evaluated by the ratio of the likelihoods of the two potentials they generate.

The Bayesian framework of Magorrian (2006) does not specifically address the problem that the orbital weights are not independent parameters of a *single, global* Schwarzschild model. Instead, the orbital weights function as (linear) coefficients for the fundamental building blocks of the model: the projected properties of the orbits which are different for each trial potential and, thus, need to be recalculated from scratch for each trial potential (cf. Appendix A). In other words, the subjects to which the orbital weights w_i refer to, *depend* on the choice of the parameters $M_\bullet, \Upsilon, i, \dots$ which have to be specified to generate the orbits. This means there is no straightforward way to define a *single* model with a common parameter space ($M_\bullet, \Upsilon, i, \dots, w_i$) as the basis for the comparison of χ^2 values obtained for different sets of trial parameters like M_\bullet, Υ or i . Instead one actually compares the goodness-of-fit of *distinct* statistical models, each with its own individual space of orbital weights w_i .

On a more fundamental level, the Schwarzschild technique is an *exact* method to find phase-space distribution functions f that obey the Collisionless Boltzmann equation:

$$\mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \phi}{\partial \mathbf{x}} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \quad (1)$$

(where ϕ is the gravitational potential and f is the phase-space distribution function of the system under study). We can formulate the Schwarzschild technique by considering a partition of phase space into a huge number of small cells. The distribution function f is represented by the large number of phase-space densities f_i in each of these small local cells. The observables of the model, such as the binned LOSVDs \mathbf{l}_{mod} of the model or its binned 3d luminosity density \mathbf{d}_{mod} , can be derived from the f_i by simple phase-space integrations, e.g. the amount of light in bin k reads

$$\mathbf{d}_{\text{mod},k} = \int_{S_k} f \, d^3r d^3v, \quad (2)$$

where the integral goes over the whole velocity space and over the subset S_k of the configuration space that represents bin k (cf. App. A for the exact definitions and vector notation of \mathbf{l}_{mod} and \mathbf{d}_{mod}). One could argue that these f_i could be used as the fundamental parameters in a *single, global* Schwarzschild model.

In fact, when we do a χ^2 minimization, we assume a statistical model for the data, specifically that each measurement $\mathbf{l}_{\text{obs},k}$ was drawn from a Gaussian distribution. The width of this distribution is commonly assumed to be known and approximated by the observational error. Only the mean, which is given by $\mathbf{l}_{\text{mod},k}$ has to be determined in the modelling process. Hence, our *statistical model* is completely determined when the \mathbf{l}_{mod} are known, which is

the case when the f_i are specified. For this, we do not even need to assume a gravitational potential or mass model. And up to this point one could indeed formulate a *single, global* dynamical phase-space model with the f_i being *independent* free parameters.

However, one issue with such a phase-space model is that arbitrary distributions of the f_i are of little interest, because most of them will be physically unrealistic or even unphysical. Besides the constraint $f_i > 0$ (positive phase-space density) another crucial requirement is given by the fact that we are only interested in equilibrium solutions, or more specifically, in solutions that represent a stationary state. The Jeans Theorem then implies that two phase-space cells i and j that happen to be located along the same orbit necessarily have to carry the same $f_i = f_j$. It is this constraint – that the phase-space density is constant along orbits – which gives equilibrium solutions of equation (1) the form of orbit superpositions and allows us to use orbits as building blocks for the models. In practice, thus, the Jeans Theorem introduces many, many nonlinear equality constraint equations for our statistical model if we formulate it on the fundamental level of the f_i . These equations require the assumption of a gravitational potential and will change as soon as the parameters of the assumed mass distribution (like M_\bullet etc.) will change: the equilibrium distributions depend on the mass structure. Hence, the set of feasible points, i.e. the set of f_i that fulfil the constraint equations will change as well. This shows that the selection parameters are not free parameters in the same sense as the f_i because they are required to define the constraint equations that regulate the domain of the model. In fact, when these constraint equations change, one effectively deals with a different statistical model with a potentially different model complexity.

The above chain of arguments can be applied to any parameter that affects the gravitational potential and, thus, the stellar orbits, e.g. the stellar Υ or the DM halo parameters. It also holds for the inclination i , or more generally, for the viewing angles. One reason is that assuming different viewing angles changes the deprojection and with it the stellar contribution to the gravitational potential. Even if the latter is kept fixed, however, the (self-consistency) constraints $\mathbf{d}_{\text{data}} = \mathbf{d}_{\text{mod}}$ imposed on the model’s stellar density do change (cf. App. A). This, in turn, means that the set of feasible points changes with the same consequences as discussed above.

Taking all this together it is clear that parameters like M_\bullet , Υ , i and the DM halo parameters on the one side and the orbital weights w_i on the other, do not ‘operate’ on the same level. It is conceptually helpful to categorize them into two different parameter classes: selection parameters and minimization parameters. The selection parameters define a *family of models* $\mathcal{F}(M_\bullet, \Upsilon, i, \dots)$. Each specific set of selection parameters defines a member of this family, i.e. a single statistical model for the measured LOSVDs. The *free parameters* of such a member model, i.e. the parameters to be minimized, are the orbital weights w_i associated with the specific set of selection parameters. From one model to another the w_i are assigned to different sets of orbits and they cannot be easily interchanged with each other across different models.

As illustrated above treating the problem of finding the optimum selection parameters as a selection of multiple distinct candidate models is a straightforward approach, whereas a parameter estimation approach using $\chi^2(M_\bullet, \Upsilon, i, \dots, w_i)$ of a single global Schwarzschild model is rather ambiguous due to the dependence of the orbital weights on the selection parameters. Furthermore the model selection framework is nothing but a generalization of the parameter estimation via χ^2 which would simplify to the latter if the multiple (sub-)models would stem from a single, statistical model (cf. Section 4). Similarly, the generalization of the Bayesian framework

introduced by Magorrian (2006) to a (Bayesian) model selection framework is straight forward and model selection criteria such as the Akaike criterion can be derived from it by choosing a suitable prior which penalizes flexibility (cf. Burnham, Anderson & Burnham 2002).

2.2 Identifying the optimum minimization parameters

Finding the optimal set of minimization parameters (i.e. orbital weights w_i) for a given set of selection parameters is in principle a direct parameter estimation problem. The goal is to minimize

$$\chi^2 = \sum_j^{N_{\text{kin}} \times N_{\text{vel}}} \left(\frac{l_{\text{obs},j} - l_{\text{mod},j}}{\Delta l_{\text{obs},j}} \right)^2. \quad (3)$$

where N_{kin} is the number of spatially resolved bins and N_{vel} the number of velocity bins in a single spatial bin. However, in Schwarzschild models one is often faced with the problem that the free parameters w_i (i.e. the number of orbits calculated for a given trial mass distribution) outnumber the observational constraints $l_{\text{obs},k}$. The related optimization problem is therefore underconstrained. In order to be able to decide on a set of unique w_i and simultaneously prevent overfitting it is common practice to introduce regularization in the fitting process. We do this by maximizing the entropy-like term proposed by Richstone & Tremaine (1988):

$$\hat{S} = S - \alpha \chi^2 \quad (4)$$

where α is a regularization parameter and S is the Boltzmann entropy defined by:

$$S = - \sum_i^M w_i \cdot \ln \left(\frac{w_i}{V_i} \right). \quad (5)$$

Here, V_i is the phase-space volume of orbit i such that w_i/V_i is the phase-space density along the orbit. The entropy term in equation (4) guarantees that the optimization problem has a unique solution for the orbital weights w (cf. the extensive discussion in Neureiter et al. 2021). The choice of equation (5) as the smoothing function is somewhat arbitrary though. The motivation behind maximizing equation (5) is that it yields $w_i \sim V_i$ and, hence, $f_i \approx \text{const}$ (in the absence of other constraints). In other words, it tends to smooth the corresponding physical phase-space distribution function rather than the distribution of the w_i , which are only parameters. Neureiter et al. (2021) discuss how the more general form

$$S = - \sum_i^M w_i \cdot \ln \left(\frac{w_i}{\omega_i} \right) \quad (6)$$

can be used to explore the full range of all (possibly degenerate) solutions for a given set of kinematic observations.

The regularization parameter α determines the smoothness of the model distribution function, meaning a small α implies the entropy term in equation (4) dominates and the phase-space density is smooth, while a larger α leads to a better fit but increases the possibility of overfitting. This raises the question which amount of regularization is optimal to represent the phase-space density of the observed galaxy. Using Monte Carlo simulations of isotropic rotator models Thomas et al. (2005) describes an approach to estimate the optimum α for NGC 4807. However, such simulations would be required for every newly modelled galaxy as the optimum α depends on the obtained data and the underlying galaxy structure. In Section 7 we will come back to this issue and provide a method to determine a more optimal amount of smoothing by treating α as a selection parameter. In

general, all the intrinsic parameters that control the behaviour of the models can be treated as selection parameters. Although we will not discuss it in full depth in this paper, another important selection parameter besides α is the number of orbits used in the models.

In conclusion, the modelling process can be described using two fundamentally different parameter layers that together define the properties of an orbit model. For the primary parameter layer the modeller has to choose a set of selection parameters, including mass and library parameters, which define the fundamental building blocks of a Schwarzschild model. In contrast, the minimization parameters that form the second layer, i.e. the orbital weights, are estimated using a form of parameter estimation for the given observational constraints. In our case this parameter estimation is performed by maximizing the entropy-like quantity in equation (4). This estimation process of the minimization parameters is only concerned with finding a good set of orbital weights. In fact, it is only possible if all selection parameters required for the calculation of the orbits are fixed and established, i.e. if a specific model out of the family of all possible models has been selected. Consequently one has to construct multiple trial models, if the goal is to optimize the selection parameters. To this end one has to sample the selection parameter space, estimate the best set of w_i for each of the resulting orbits (using equation 4), and compare the properties of each model afterwards. Since these models consist of different building blocks they are fundamentally distinct and one has to establish a framework to evaluate which one of them results in the best approximation of the galaxy's underlying structure.

As already stated above, the common approach to evaluate different models \mathcal{F}_A and \mathcal{F}_B with different selection parameters M_* , Υ , ... based on $\Delta\chi^2 = \chi_A^2 - \chi_B^2$ (cf. equation 3) is not optimal. We will show that judging models solely in terms of their goodness-of-fit can induce biases in the recovered selection parameters. The next section demonstrates such a case for a selection parameter that is very heavily impacted by this: The inclination of axisymmetric models.

2.3 Biases in selection parameters: the inclination as an example

We tested the Schwarzschild technique described in the previous section by applying it to a simulated spherical galaxy. For this purpose we created a number of mock LOSVDs obtained from an N -body model of an isotropic spherical galaxy (Hernquist 1990). We sampled the sphere with $N = 10^9$ particles and added Monte Carlo noise to the corresponding LOSVDs. The underlying Hernquist density we used had a total mass $M = 50 \times 10^{10} M_\odot$, a mass-to-light ratio $\Upsilon = 1$ and an effective radius of 10kpc. The stellar kinematics of the N -body was projected to a grid of 80 angular and radial bins reaching out to approximately 2.3 effective radii. In total we simulated 10 independent observations of this spherical galaxy by adding random uncorrelated Gaussian noise with a standard deviation of two per cent of the maximum LOSVD value in the respective spatial bin. Quantifying this noise in terms of the Gauss–Hermite coefficients (cf. van der Marel & Franx 1993) this translates to an average error of $\Delta\sigma/\sigma \approx 2$ per cent and $\Delta h_4 \approx 0.02$. The resulting Gauss–Hermite coefficients of one of these mocks are shown in Figs B1 and B2 (Appendix B). Without the Gaussian noise the N -body models follow the semi-analytic Gauss–Hermite profiles of (cf. Baes, Dejonghe & Buyle 2005) very closely, which means that the odd Gauss–Hermite coefficients are zero everywhere, the velocity dispersion σ has a maximum at intermediate radii and h_4 is slightly negative but increases significantly for $r \rightarrow 0$.

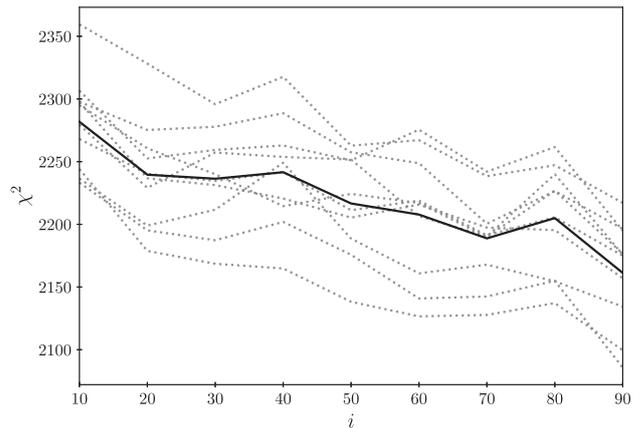


Figure 1. Axisymmetric Schwarzschild fits to a spherical N -body simulation. Dashed, grey lines represent fits to different mock data sets obtained by adding random noise to the same original spherical model kinematic. The solid line is the corresponding arithmetic mean. The number of kinematic data points is $N_{\text{data}} = 2640$. Even though the input model is spherical, the χ^2 varies as a function of the inclination assumed in the fit.

We modelled the spherical mocks with differently inclined, axisymmetric Schwarzschild models using the same spherical Hernquist density that was used to create the N -body in the first place. Thus, effectively we used the same orbit library for all models, yet the orbits were projected under different viewing angles. The resulting χ^2 values as a function of the inclination are shown in Fig. 1.

All 10 mocks exhibit a significant bias towards edge-on models, implying that the ability of an orbit model to fit data depends on the angle it is observed at. The reason for this behaviour is the axisymmetry of our Schwarzschild models: Every orbit in our model exists in a prograde and a retrograde version. In the case of an edge-on model the LOSVD of the prograde and retrograde versions have opposite signs and are each a unique contribution to the model's LOSVDs in equation (A1). This is contrasted by a face-on model where the prograde and retrograde versions are identical in projection, thus effectively reducing the number of unique base functions in equation (A1). This weakens the flexibility of our model LOSVDs to fit the observations and consequently leads to a higher χ^2 of face-on models when compared to their edge-on counterparts. Since this effect is intrinsic to the modelling technique, the edge-on bias will also be present when non-spherical galaxies are modelled, thus diminishing the possibility of successfully constraining the actual inclination of a galaxy. Therefore we should aim to quantify the variable model flexibility and use that to correct our results for the inclination constraints.

3 MODEL FLEXIBILITY

In Section 2 we showed that evaluating Schwarzschild models based on their χ^2 can suffer from a bias if models with varying flexibility are compared. Therefore we introduce two methods in Section 3.1 with the goal to estimate this flexibility, followed by an outline of three model selection approaches in Section 4 which exploit this information to improve the constraining power of our models.

3.1 Estimating the model flexibility in Schwarzschild models

In statistical modelling a model's general ability to fit data is quantified by its number of free parameters m , because every free

parameter reduces the degrees of freedom of the χ^2 distribution. This means if we assume uncorrelated Gaussian noise and quantify the deviations of N observed data points to a statistical model with m free parameters then the expectation value of the respective χ^2 will be $E(\chi^2) = N - m$ with variance $\text{Var}(\chi^2) = 2(N - m)$. Therefore it is only appropriate to compare models using χ^2 if they have the *same number of free parameters*.

For linear models without prior constraints the number of free parameters is a well-defined property that can be quantified using a linear algebra framework (e.g. Hastie, Tibshirani & Friedman 2013). However, our Schwarzschild models are nonlinear due to the entropy term introduced in the estimation process of the orbital weights (equation 4). But even if they were linear the estimation of the number of free parameters would be a non-trivial task because we demand the orbital weights w_i to be non-negative. This requirement is a prior constraint that limits the accessible parameter space and thus reduces the flexibility of the models in an unpredictable fashion (e.g. Andrae et al. 2010). This is because the amount by which priors reduce the flexibility of a model depends on the data that are getting fitted, i.e. in this case on the structure of the galaxy under study and on the noise pattern of the data. Therefore we need a more generalized approach to quantify the actual flexibility of our Schwarzschild models, like the definition introduced by Ye (1998) which extends to nonlinear models (see details below).

We call the quantity describing this flexibility the number of effective free parameters m_{eff} . Both, the nonlinearity and the non-negativity prior affect the flexibility of the Schwarzschild models in a complicated fashion. This prevents a direct calculation of the model flexibility and we rely on bootstrap iterations to estimate m_{eff} . After having fitted a Schwarzschild model to the observed LOSVDs we start a number of K bootstrap iterations. In each iteration we add random Gaussian noise to the original fit \mathbf{l}_{fit} . The standard deviation of this artificially added noise is based on the observational error $\Delta \mathbf{l}$. When \mathbf{l}_{fit} is a statistically ‘good’ fit, i.e. if it can quantitatively explain the actual data in line with the estimated observational errors, then the new bootstrap data $\mathbf{l}_{\text{bootdata}}$ should mimic the observed data: each bootstrap data point is redrawn from the (assumed) same distribution as the corresponding real data point. If the model is not a good fit (e.g. if the mass parameters are completely wrong) this assumption is not valid, fortunately such models can be discarded easily and are of no further interest. Given this bootstrap assumption (i.e. the bootstrap samples are representative of the observed data), we can estimate the flexibility of the original fit model by fitting each of the K bootstrap data sets with the same selection parameters that were used for the original fit. We denote bootstrap fits as $\mathbf{l}_{\text{bootfit}}$. We tested two independent methods to estimate m_{eff} based on these bootstrap fits. The first one measures the reduction of χ^2 directly by calculating the χ^2 before the bootstrap fit

$$\chi_{\text{prior}}^2 = \sum_{j=1}^N \left(\frac{\mathbf{l}_{\text{bootdata},j} - \mathbf{l}_{\text{fit},j}}{\Delta \mathbf{l}_j} \right)^2 \quad (7)$$

and after it:

$$\chi_{\text{posterior}}^2 = \sum_{j=1}^N \left(\frac{\mathbf{l}_{\text{bootdata},j} - \mathbf{l}_{\text{bootfit},j}}{\Delta \mathbf{l}_j} \right)^2. \quad (8)$$

We can then exploit the expectation values $E(\chi_{\text{prior}}^2) = N$ and $E(\chi_{\text{posterior}}^2) = N - m_{\text{eff}}$ to estimate m_{eff} by averaging over all boot-

strap iterations K :

$$m_{\text{eff}} = \frac{1}{K} \sum_{k=1}^K m_k = \frac{1}{K} \sum_{k=1}^K (\chi_{\text{prior},k}^2 - \chi_{\text{posterior},k}^2). \quad (9)$$

An alternative approach to estimating m_{eff} is to calculate:

$$m_{\text{eff}} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^N \frac{1}{\Delta \mathbf{l}_i^2} (\mathbf{l}_{\text{bootfit},k,i} - \mathbf{l}_{\text{fit},i}) (\mathbf{l}_{\text{bootdata},k,i} - \mathbf{l}_{\text{fit},i}). \quad (10)$$

If we assume that the expectation values of the bootstrap fit and data are given by $E(\mathbf{l}_{\text{bootfit},i}) = E(\mathbf{l}_{\text{bootdata},k,i}) = \mathbf{l}_{\text{fit},i}$, equation (10) equals the sum of normalized covariances:

$$m_{\text{eff}} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^N \frac{\text{cov}(\mathbf{l}_{\text{bootfit},k,i}, \mathbf{l}_{\text{bootdata},k,i})}{\Delta \mathbf{l}_i^2}. \quad (11)$$

In this form m_{eff} is equivalent to the concept of generalized degrees of freedom for nonlinear models developed by Ye (1998). The above bootstrap approach is versatile and can be adopted to estimate the flexibility of very complex statistical models such as orbit superposition models. However, it comes at the cost of computational performance as it requires several additional model fits, i.e. identifying the optimum orbital weights (cf. Section 2.2) for each of the constructed bootstrap data sets. Fortunately, though, the bootstrap iterations do not require additional orbit integrations as one can reuse the orbit library of the original model fit.

As a first proof-of-concept we applied our bootstrap methods to simple polynomial function models where the number of free parameters (the number of polynomial coefficients to be recovered) can be counted. We found that equations (9) and (10) are equivalent within the adopted numerical precision: both can be used to calculate the number of free parameters correctly. For the more complex Schwarzschild models we found a slight offset, which is likely because the condition $E(f_i) = \mathbf{l}_{\text{fit},i}$ is not fulfilled for all data points when the orbital weights are forced to be non-negative but the bootstrap noise is assumed to follow a Gaussian distribution (which sometimes implies negative LOSVD values). Fortunately, this offset is approximately constant thus our results do not depend on the chosen estimation approach. For the rest of the paper, we use the covariance approach (equation 10). We applied both estimates to the modelling of the spherical toy galaxy introduced in Section 2.3 using $K = 50$ bootstrap iterations. The resulting m_{eff} as a function of the inclination are depicted in Fig. 2. We found that the behaviour of χ^2 is mirrored by the number of effective parameters m_{eff} : where the χ^2 gets lower, m_{eff} increases. In fact, the χ^2 variation in Fig. 1 can be entirely explained by m_{eff} . This supports our hypothesis that the apparent edge-on bias is caused by the varying flexibility of the axisymmetric models with the inclination. Using the above bootstrap estimation methods we also identified a multitude of other factors influencing the general ability of a Schwarzschild model to fit a given set of data. The most dominant other selection parameter is the regularization value α . In Fig. 3 m_{eff} is shown as a function of α for three edge-on orbit libraries with different number of orbits N_{orbit} . All three libraries are constructed with the correct gravitational potential that was used to create the mock observation they attempt to fit. A decrease in α suppresses the freedom of the orbital weights and, thus, restricts the model’s flexibility. In the limit of $\alpha \rightarrow 0$ the model becomes completely rigid resulting in $\chi^2 \approx N$. For very large values of α both, χ^2 and m_{eff} , appear to converge to a constant value. We will come back to the regularization in Section 7. An increase in orbits typically leads to a (nonlinear) increase in

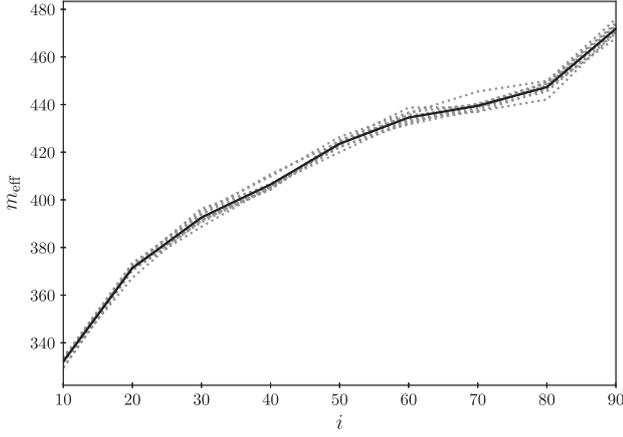


Figure 2. The number of effective parameters m_{eff} as estimated by equation (10) of the models shown in Fig 1. The unexpected χ^2 variation in Fig. 1 can be explained by the varying model flexibility.

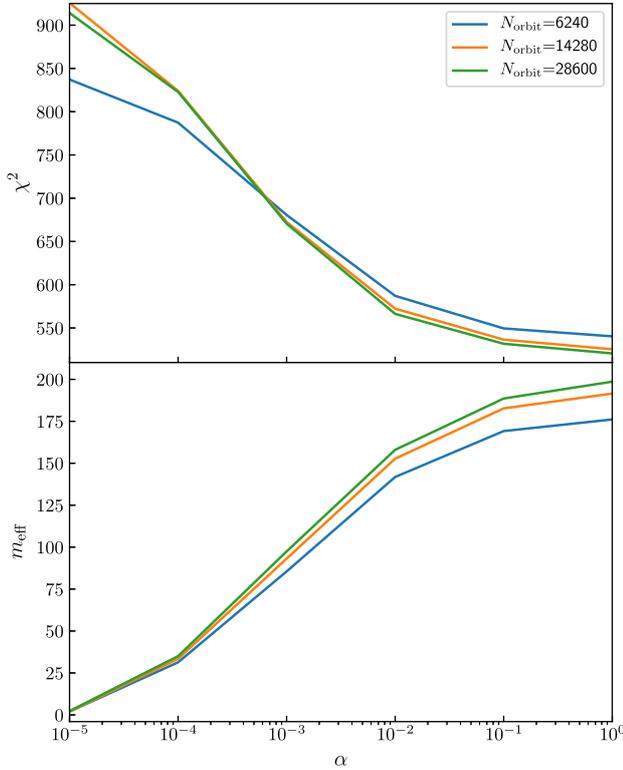


Figure 3. Top panel: χ^2 as a function of the regularization parameter α for three orbit libraries with different number of orbits N_{orbit} . A set of 782 kinematic data points of a single mock of an edge-on toy galaxy is being fitted by the orbit models. All orbit models are viewed edge-on and employ the correct density and gravitational potential that was used to create the mock data in the first place. Bottom panel: The same as the top panel but for the number of effective parameters m_{eff} . More orbits and less regularization lead to an increase in the intrinsic model flexibility, resulting in correspondingly better fits.

model flexibility, resulting in correspondingly smaller χ^2 values. However, for very small α the orbit library with only 6240 orbits achieves a better χ^2 even though its intrinsic flexibility m_{eff} is consistently lower than that of the other libraries. This is because all orbit models are very rigid for such small values of α and the

kinematic data of the toy galaxy shown in Fig. 3 was generated using a maximum entropy Schwarzschild model with exactly $N_{\text{orbit}} = 6240$. We will describe the construction of such mocks in more detail in Section 5.1.

4 SELECTION PARAMETER OPTIMIZATION VIA MODEL SELECTION

The effective parameters quantify the model flexibility that is synonymous with the size of the parameter subspace accessible to the orbital weights. In principle any selection parameter can influence this freedom, meaning models with different sets of selection parameters should not be compared without taking into account their actual flexibility. While we showed how we can estimate this flexibility m_{eff} in the last Section 3 the question remaining is how we can use this information to choose the model with the best set of selection parameters. Since this means we need to choose the ‘best’ model out of a pool of models with different flexibilities the only option is to work within a model selection framework. We tested three different model selection approaches. The first and what we call intuitive approach is based on the following assumption: if we want to treat all models a priori equivalent then we should demand that the expectation value of our evaluation statistic should be identical for all models. Therefore we should minimize:

$$\chi^2 + m_{\text{eff}} \quad (12)$$

because $E(\chi^2 + m) = N_{\text{data}}$ holds for all models that are able to emulate the observed data. The second, information based approach is to maximize the Akaike information criterion (AIC) or equivalently to minimize:

$$\chi^2 + 2m_{\text{eff}}. \quad (13)$$

This means that the AIC approach penalizes flexible models more than the intuitive approach in equation (12) does. Since both of the above approaches differ only by the relative importance of the model flexibility m_{eff} , our third, more generalized selection approach is to minimize:

$$\chi^2 + w_m m_{\text{eff}} \quad (14)$$

where the factor w_m is a free parameter calibrated using a set of simulations described in the following Section 5. Equation (14) includes the intuitive approach ($w_m = 1$) and the AIC approach ($w_m = 2$) as special cases.

5 APPLICATION: INCLINATION RECOVERY IN SIMULATED GALAXIES

We created a number of toy galaxies with different inclinations to test the model selection framework proposed in the previous section. The goal was to recover their true inclination. Analytical distribution functions for non-spherical models are mostly not very realistic in terms of the orbital anisotropy they imply. Consequently we decided to create the noise-free LOSVDs of our toy galaxies on the basis of Schwarzschild models in realistically flattened 3d mass distributions. This approach is very flexible as modifications to the entropy term can be used to generate almost any desired orbital anisotropy (cf. Neureiter et al. 2021). We cross-checked the validity of this toy model generation on the Hernquist sphere: the modelling results obtained using either LOSVDs from the N -body-particle sampled analytic distribution function (Section 2.3) or from suitable Schwarzschild models in the Hernquist potential did not show any significant differences. Using this Schwarzschild

Table 1. Table of tested toy galaxies with different intrinsic axis ratios q , angular momentum biases λ , and inclinations i . Beyond that, we varied the velocity resolutions Δv_{los} of the LOSVDs for some models. We used three different spatial grids for the projected kinematics of these galaxies that are typical for modern observations with IFU spectrographs. The wide field FoV covers at least the kinematics within the galaxy’s effective radius. The other two grids have smaller FoVs; however, even the smallest field comfortably covers a hypothetical sphere of influence of a black hole typical for the size of such a toy galaxy. Furthermore, all three spatial grids have increased resolution in the centre where such a sphere influence would be located.

Name	Density	Axis ratio q	Angular momentum bias λ	Inclination	Δv_{los} (km s ⁻¹)	Spatial grid
Galaxy A	Early-type	0.6	0.0	60°	76	Wide field
Galaxy B	Early-type	0.4	0.0	70°	76	Wide field
Galaxy C	Early-type	0.6	0.0	45°	76	Wide field
Galaxy D	Early-type	0.6	0.5	60°	76	Wide field
Galaxy E	Early-type	0.6	1.0	60°	76	Wide field
Galaxy F	Early-type	0.6	0.0	60°	65	Wide field
Galaxy G	Early-type	0.6	0.0	60°	54	Wide field
Galaxy H	Late-type	~0.2	1.0	55°	30	Intermediate field
Galaxy I	Late-type	~0.2	0.0	55°	30	Intermediate field
Galaxy J	Late-type	~0.2	1.0	55°	30	Small field
Galaxy K	Late-type	~0.2	0.0	55°	30	Small field

model construction we simulated the kinematics of both, late-type and early-type galaxies. The former with maximum-entropy models of a realistic mass distribution obtained by deprojecting the surface brightness profile of a real disc galaxy and the latter with models of a flattened Hernquist distribution. Table 1 is a comprehensive list of all toy galaxies we investigated and their most relevant characteristics. To create realistic mock observations of these toy galaxies the stellar kinematics of the resulting orbit model must be projected to the sky. To this end we projected the kinematics to spatial grids borrowed from real SINFONI and MUSE observations (e.g. Mehrgan et al. 2019) and convolved them with the corresponding seeing. For the early-type galaxies we used a Voronoi grid that covers a large field of view (FoV) and extends well beyond the toy galaxy’s effective radius. To project the late-type galaxies we used regular grids with a smaller FoV but higher spatial resolution. The FoV is still large enough to cover the sphere of influence of a hypothetical black hole with a mass typical for the respective galaxy. We simulated 10 independent observations of each toy galaxy by adding random uncorrelated Gaussian noise with a standard deviation of two per cent of the maximum LOSVD value in the respective spatial bin (cf. Section 2.3). Since we generated rotation in some of the toy galaxies (see below) not all of them intrinsically represent a maximum-entropy state.

After the addition of noise, all resulting mock observations were modelled with Schwarzschild models using 6240 orbits and a regularization parameter of $\alpha = 1.67$. For our toy galaxies this regularization value was large enough to ensure that $\chi^2(\alpha)$ has roughly converged to a constant value, in other words the flexibility of the respective models has plateaued and a further increase in α does not improve the goodness-of-fit significantly. To recover the mock galaxies’ inclinations, we had to model them under different assumed viewing angles. To do so, we projected the true intrinsic density of each toy galaxy on the sky and deprojected the resulting mock images assuming the various inclinations probed by the dynamical models. For these deprojections we used the Metropolis-Algorithm of Magorrian (1999). It is usually not required to probe every viewing angle as not all inclinations are necessarily compatible with a given photometry. In axisymmetric systems in particular the observed flattening on the sky determines a minimum possible inclination that corresponds to an infinitely thin distribution. But even for triaxial galaxies the range of plausible viewing angles can be narrowed down (cf. de Nicola et al. 2020).

5.1 Simulated galaxies

The early-type toy galaxies have a mass-to-light ratio $\Upsilon = 1$ and are based on the flattened density profile:

$$\rho(a) = \frac{M}{2\pi} \frac{a_{\text{scale}}}{q} \frac{1}{a(a + a_{\text{scale}})^3} \quad (15)$$

which is similar to the Hernquist sphere (cf. Hernquist 1990). The variable a parametrizes the spheroidal isodensity surfaces. While not all early-type toy galaxies have the same intrinsic axial ratio q , their total mass $M = 5 \times 10^{11} M_{\odot}$, effective radius $r_{\text{eff}} = 10\text{kpc}$ and distance $d = 141.8\text{Mpc}$ are always the same. The density for the late-type toy galaxies was obtained by deprojecting the photometry of a real late-type galaxy (NGC 3489, cf. Nowak et al. 2010). Beyond that we also borrowed the assumed inclination ($i = 55^\circ$), distance ($d = 12.1\text{Mpc}$) and seeing from said galaxy for the set-up of the late-type galaxies.

After establishing the density distribution for each galaxy we generated its kinematics by constructing maximum-entropy Schwarzschild models based on this density. The kinematics of the models were then projected on to the sky and convolved with realistic spatial point spread functions (PSF). The (non-parametric) LOSVDs for the early-type galaxies were projected on to a set of 34 Voronoi bins using the method of Cappellari & Copin (2003). To make the mock data as realistic as possible, this observational set-up and the respective PSFs were borrowed from real SINFONI and MUSE observations of a massive elliptical galaxy (ESO 325-G004). The mock SINFONI observations cover the inner 1 arcsec with high enough spatial resolution to resolve a central supermassive black hole while the corresponding MUSE observations encompass the larger-scale kinematics of the galaxy out to about 2 half-light radii. The typical size of these Voronoi bins varies with radius from about $\sim 0.1\text{ kpc}$ (or $\sim 0.17\text{ arcsec}$) in the centre to $\sim 3\text{ kpc}$ (or $\sim 4\text{ arcsec}$) in the outer bins. An example Gauss–Hermite map for the MUSE grid of one of the resulting mock observations is shown in Fig. B3 (Appendix B). For the projection of the Schwarzschild kinematics of late-type galaxies we tried two regular grids with differently sized FoVs (cf. Table 1) resulting in 30 and 70 spatial bins respectively. For the PSF convolution we borrowed the observed PSFs of the same late-type galaxy NGC 3489 that we used to create the intrinsic density distribution.

All maximum entropy galaxies display characteristic features of non-rotating flattened systems such as smaller observed velocity

dispersion σ and h_4 on the minor axis, but overall they behave similar to the spherical toy galaxy of Section 2.3 in that their LOSVDs are symmetric and have no net rotation. Since this is not representative of most real galaxies (cf. Emsellem et al. 2011), we also modelled rotating toy galaxies with the goal to investigate how this affects the inclination recovery. We introduced rotation by exploiting the symmetry of our axisymmetric orbit libraries: each orbit of our libraries exists in a prograde and a retrograde version. In a maximum-entropy model, the weights of the two orbits in a prograde–retrograde pair, denoted by w_i^+ and w_i^- , are identical and consequently the orbits cancel each other’s rotation signal. We can simply manipulate this prograde–retrograde balance to create a toy galaxy with net rotation by reassigning new orbital weights:

$$\begin{aligned}\hat{w}_i^+ &= \frac{1+\lambda}{2} \cdot (w_i^+ + w_i^-), \\ \hat{w}_i^- &= \frac{1-\lambda}{2} \cdot (w_i^+ + w_i^-)\end{aligned}\quad (16)$$

where the free parameter $\lambda \in [-1, 1]$ biases the total angular momentum L_z along the library’s symmetry axis. For $\lambda = 1$, only prograde orbits contribute to the observed kinematics, resulting in a maximum positive rotation signal, while $\lambda = -1$ implies a maximum negative rotation signal as only retrograde orbits are populated. We can also recover our original non-rotating, maximum entropy model by choosing $\lambda = 0$. Using this weight manipulation we created two rotating mock galaxies (see Table 1) with an angular momentum bias of $\lambda = 1$ (Galaxy E) and $\lambda = 0.5$ (Galaxy D). In addition to the intrinsic axial ratio q , the inclination i and the angular momentum bias of our toy galaxies we also experimented with a change of the velocity binning Δv_{los} by altering the maximum velocity v_{max} of the LOSVDs while keeping the number of velocity bins n_{vel} constant.

Qualitatively all simulated galaxies showed similar behaviour when modelled under different assumed inclinations. Therefore we first aim to outline this common behaviour using Galaxy D as a generic case. This galaxy has an axial ratio $q = 0.6$ and was originally projected at an angle $i = 60^\circ$, resulting in an apparent axial ratio $q' \approx 0.72$. In axial symmetry then, inclinations $i \lesssim 44^\circ$ are not compatible with the projected surface brightness and do not have to be considered. When modelling the galaxy, we sampled the inclination parameter linearly in the interval $[50^\circ, 90^\circ]$ with a step size of $\Delta i = 10^\circ$.

Despite the fact that the galaxy is fairly inclined with a true inclination $i = 60^\circ$, the χ^2 of the Schwarzschild models suggest that the edge-on model is the best fit on average (Fig. 4). This undesired behaviour mirrors the situation in the Hernquist sphere discussed in Section 2.3. It is caused by a drastic change in the model’s flexibility with inclination, as can be seen in Fig. 5. The figure shows that the number of effective parameters increases rapidly as a function of inclination. Thus, the lower χ^2 at higher inclination are not necessarily an indication that the fits to the data are intrinsically better at larger i . Rather, the χ^2 gets lower because the model is more flexible. In other words, the measure for which χ^2 represents a ‘good’ fit has shifted towards lower values. In fact, if we take the varying model flexibility into account and compare the models within our model selection framework, then the inclination bias disappears (Fig. 6). For both, the intuitive and the AIC approach, we can recover the galaxy’s true inclination on average very well. If any, then the fits to the individual mock realizations seem to scatter slightly above the true inclination for the intuitive approach and slightly below the true inclination for the AIC approach, but on average any related bias is very small, $\Delta i \sim 5^\circ$. Still, the distribution of the individual fits suggests that even better results might be achieved with some

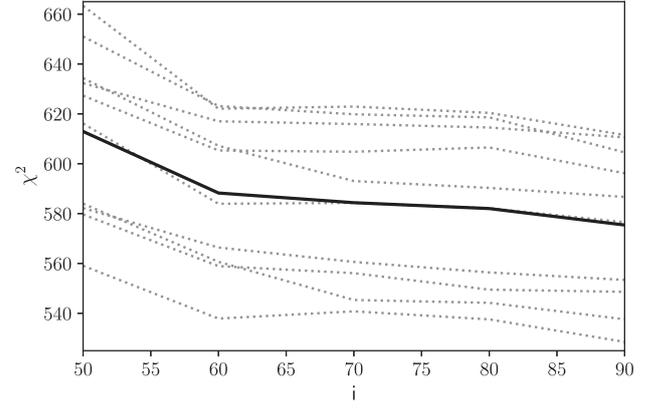


Figure 4. χ^2 as a function of the assumed library inclination when modelling toy galaxy D. Dotted lines: χ^2 for the individual mocks. Solid line: The arithmetic mean χ^2 of all 10 mocks. The true inclination of the galaxy is $i = 60^\circ$. All models had 6240 orbits and a fixed regularization parameter $\alpha = 1.67$.

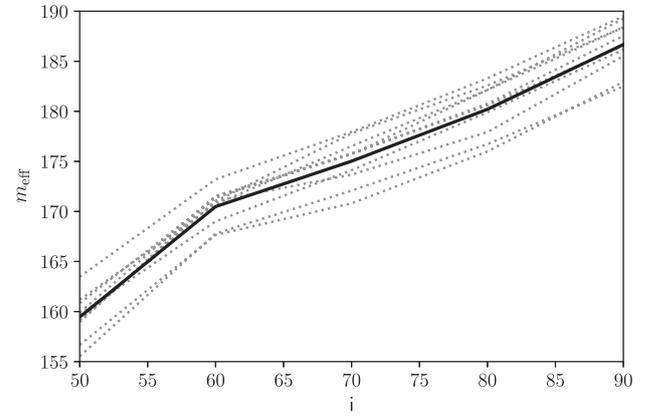


Figure 5. As Fig. 4 but for the number of effective parameters n_{eff} of the model fits to the mocks of Galaxy D that are shown in Fig. 4.

optimized $w_m \in [1, 2]$. In fact, we repeated this analysis for all toy galaxies of Table 1. For a representative subset of these simulated galaxies, Fig. 7 shows the recovered inclination again averaged over fits to ten mock data realizations (solid black lines) as a function of w_m . The figure shows that the behaviour of the fits to Galaxy D described above is actually generic. For $w_m = 0$ (simple χ^2 minimization) we find that the ‘best-fit’ inclination is almost always $i = 90^\circ$ or close to it. With increasing w_m , the influence of the varying model flexibility on the recovered inclination becomes stronger and the ‘best-fit’ inclination moves away from $i = 90^\circ$.

For most early-type galaxies, the true inclination is very well recovered with $w_m \approx 1.5$, while larger w_m lead to a slight bias towards too small inclinations. Similarly, the inclination recovery for the late-type galaxies is improved when selecting models with $w_m > 1$, however, optimum results are achieved in the Akaike regime ($w_m = 2$). A bias towards too low inclinations is strongly suppressed for the late-type toy galaxies as inclination angles significantly smaller than the true $i = 55^\circ$ are excluded on the basis of the apparent flattening in the photometric data.

Given the modelling results of all toy galaxies in Table 1 we conclude that a simple χ^2 minimization for the *selection parameter* i leads to a very significant bias. This is caused by the fact that edge-on models are much more flexible than models with a lower inclination

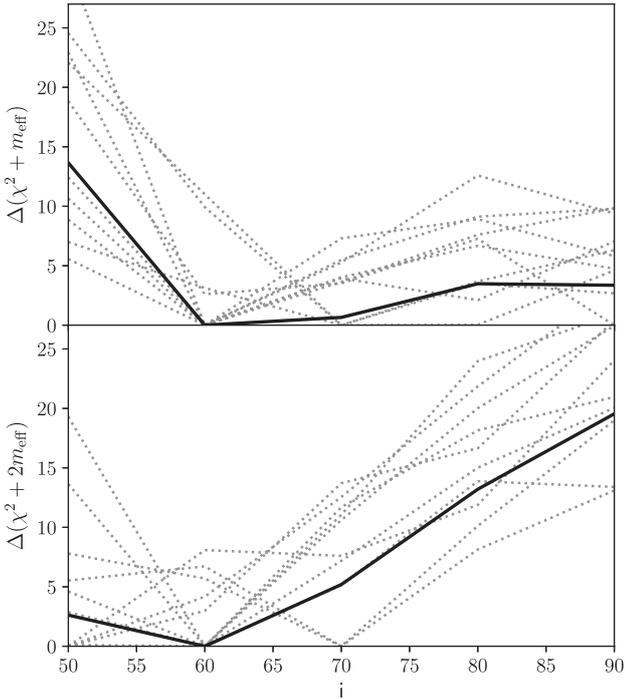


Figure 6. Similar to Fig. 4 but for a model selection framework. The absolute minimum of $\chi^2 + w_m m_{\text{eff}}$ has been subtracted for each mock, such that only the relative differences Δ to this minimum is plotted. Top panel: Inclination recovery using the ‘intuitive’ approach $\chi^2 + m_{\text{eff}} \rightarrow \min$ (equation 12). Bottom panel: Inclination recovery using the AIC approach $\chi^2 + 2m \rightarrow \min$ (equation 13). For both approaches we find that the apparent inclination bias that results from a simple χ^2 minimization (Fig. 4) disappears when the changing model flexibility (Fig. 5) is taken into account. The true inclination of Galaxy D is $i = 60^\circ$.

and, hence, yield much smaller minimum χ^2 values. However, when taking the differences in model flexibility into account by penalizing each model with an additive term proportional to the effective number of free parameters, the inclination recovery actually works very well. With the new model selection framework, we could recover the inclination of all toy galaxies to within $\Delta i = 5^\circ$ on average.

The best results were obtained when weighting the influence of the model flexibility with a proportionality factor of $w_m \approx 1.5$, which is a value right between the ‘intuitive’ approach and the AIC. We will come back to this in Section 7 where we also include the regularization into the analysis.

5.2 Flexibility and underlying galaxy structure

A unique feature of the flexibility of nonlinear statistical models is that the number of effective parameters can generally depend on the underlying data generating process (cf. Ye 1998). In the case of our dynamical models this means that the flexibility is not a universal property of the model alone, i.e. a property of the orbit library, but instead can additionally depend on the underlying galaxy structure.

This dependence becomes especially apparent when comparing the dynamical modelling results of rotating and non-rotating galaxies. We investigated this by fitting mock data sets of a sequence of toy galaxies which were identical except for the angular momentum bias λ (equation 16). One toy galaxy was generated with $\lambda = 0.0$ and represents a non-rotating maximum-entropy galaxy (Galaxy A in Table 1). The two other toy galaxies were constructed to be rotating

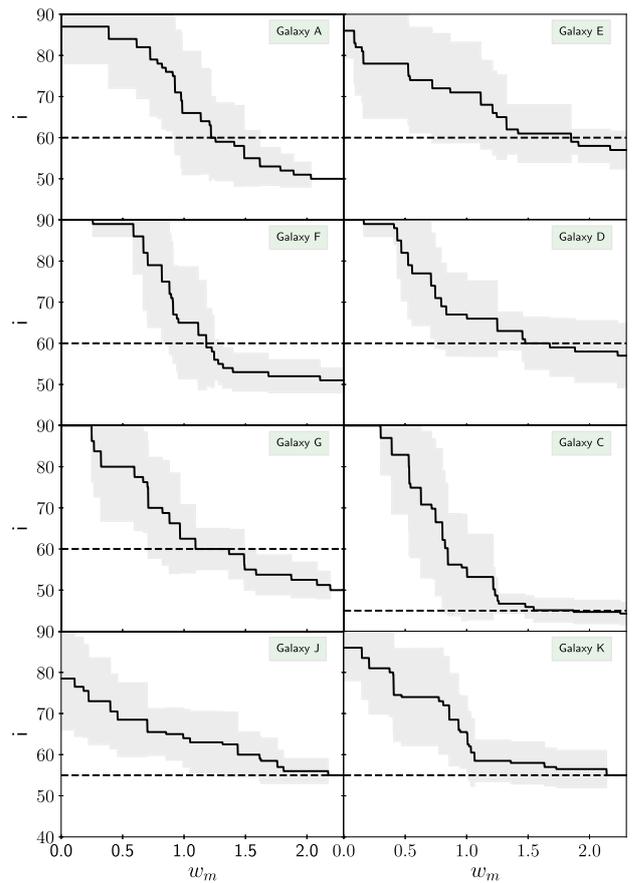


Figure 7. Inclination recovery as a function of the weight parameter w_m for some of the toy galaxies of Table 1. Solid Line: The recovered inclination averaged over fits to all 10 mock observations of each simulated galaxy. Grey area: The corresponding 1σ level estimated from the mock sample. Dashed Line: The true inclination of each toy galaxy. All simulated galaxies suffer from a dominant edge-on bias when evaluated with χ^2 ($w_m = 0$). For the early-type galaxies we find that a model selection with an intermediate $w_m \in [1.2, 1.8]$ achieves an optimum inclination recovery. For the late-type galaxies a slightly higher $w_m \rightarrow 2$ appears to be optimal. However, all these results depend on the regularization α as we will demonstrate in Section 7.

by setting $\lambda = 0.5$ (Galaxy D) and $\lambda = 1.0$ (Galaxy E). The resulting (noise-free) rotation maps of the latter two toy galaxies are shown in Fig. 8. After having added noise we analysed the three toy galaxies with an identical orbit library and estimated the flexibility of the fits using the bootstrap calculations of Section 3.1 above. We repeated this for five different inclinations from $i = 50^\circ$ to $i = 90^\circ$.

Fig. 9 shows the resulting estimated number of effective parameters. Note that for a given inclination, the mock data of the toy galaxies with different λ were fitted with exactly the same orbit library. Despite this fact, however, the model’s ability to fit the data changes with λ . On the one hand, we again see how the overall flexibility of the model increases with i . On the other hand, however, we also see that the more net rotation is present in the underlying galaxy, the less flexible the model appears. We argue that the reason for this dependency is the fact, that with increasing rotation the light in the observed LOSVDs is more and more concentrated on either positive or negative velocities. While the amount of light on one side of the LOSVD therefore increases, it decreases on the opposite side and more and more velocity bins in each LOSVD tend to carry only little or even no light when the angular momentum bias λ is strong.

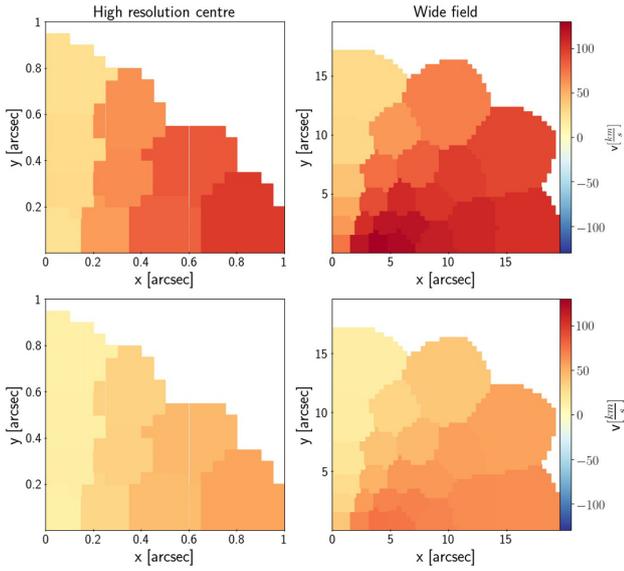


Figure 8. The (noise-free) rotation maps of two toy galaxies with angular momentum bias $\lambda = 1.0$ (top panels) and an intermediate $\lambda = 0.5$ (bottom panels). Panels on the left side show the centre with the higher-resolution data, panels on the right show the simultaneously modelled wide-field data.

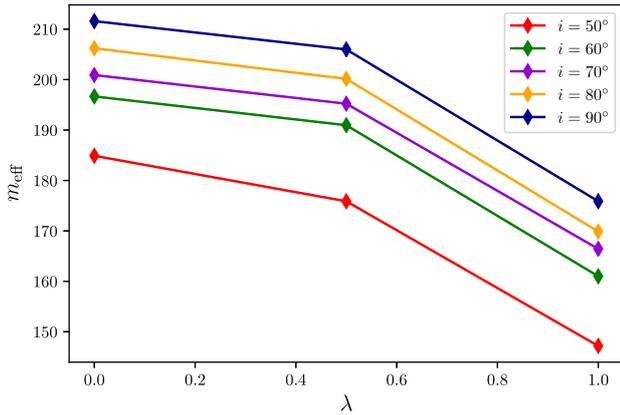


Figure 9. The number of effective parameters m_{eff} estimated for axisymmetric Schwarzschild models of rotating toy galaxies with different angular momentum bias λ . For each toy galaxy, the results are averaged over 10 mocks and the regularization parameter was set to $\alpha = 1.67$. Even though models at the same inclination i were obtained with identical orbit libraries, the number of effective parameters m_{eff} does not stay the same but decreases with the angular momentum bias. This is true at any inclination.

After the addition of Gaussian noise such (nearly) dark velocity bins can formally even carry a negative signal. Since we require the orbital weights to be positive, such formally negative LOSVD values cannot be fitted. Thus, because with more rotation more LOSVD bins are prone to carry a negative signal due to noise, the model’s flexibility as estimated by the number of effective parameters will shrink.

This correlation of m_{eff} with the rotation implies that the flexibility of Schwarzschild models depends on the underlying phase-space distribution function of the galaxy under investigation. The same orbit library can be more responsive to one data set than to another. This is in contrast to the flexibility of linear models where the number of free parameters is independent of the underlying data generating process (as long as no constraints on the free parameters are applied). We conclude that the number of effective parameters of a given orbit

model should be estimated individually for each new fit. When modelling the kinematics in the different quadrants of the same galaxy separately, for instance, even if one samples the mass parameters from the same parameter grid, the number of effective parameters needs to be estimated separately for each fit in each quadrant.

6 BIASES IN OTHER SELECTION PARAMETERS

The dependence of χ^2 on the model flexibility is most prominent for differently inclined models, however, in principle any selection parameter can suffer from biases introduced by a variable model flexibility. We investigated this by modelling toy galaxy E with different model inclinations and mass-to-light ratios Υ .

The result is displayed in the top two panels of Fig. 10, which show contours of χ^2 and m_{eff} as a function of the mass-to-light ratio Υ and the inclination i , averaged over 10 mocks. In the direction of the inclination we observe the edge-on bias established in the previous sections. However, the contours reveal an additional bias along the mass-to-light ratio axis. A simple χ^2 -minimization does not result in a minimum at the true mass-to-light ratio $\Upsilon_{\text{true}} = 1.0$ but instead is biased towards slightly larger ratios. The fact that the χ^2 behaviour is mirrored by the number of effective parameters m_{eff} suggests that this increase in model flexibility causes the bias. While this bias of the mass-to-light ratio is present for all modelled inclinations, it appears to be most significant towards $i = 90^\circ$. Thus, a model evaluation based on a simple χ^2 -analysis would not only overestimate the inclination but would also misjudge the total mass of the system.

We argue that the reason for this Υ - m_{eff} -correlation is due to the higher escape velocity in models with larger total mass. Orbits with a higher escape velocity can potentially occupy velocity bins that are unattainable at lower masses. Models at higher masses will therefore be able to fit more LOSVD bins (if the LOSVDs are sampled out to sufficiently large velocities). This means that they have an increased model flexibility when compared to their lower Υ counterparts.

If we take this effect into account and judge the dynamical models within the model selection framework, then the recovery of the original mass-to-light ratio and inclination becomes highly accurate. This is shown in the bottom panel of Fig. 10 which shows the constraints obtained with the calibrated model selection approach (i.e. $w_m = 1.5$ equation 14). Both inclination and mass-to-light ratio biases have disappeared and the constraints are now tightly centred around the correct model with 60° model and $\Upsilon = 1$. Similar to the inclination an intermediate calibration weight of $w_m \approx 1.5$ proves to be most successful in recovering the correct mass-to-light ratio. The intuitive and the Akaike approach also manage to select the correct model on average, however, their contours are slightly skewed towards larger Υ for $w_m = 1$ and smaller Υ for $w_m = 2$.

The orbit models shown in this and the previous Section 5 were modelled with a fixed regularization α . However, the constraining power of the model selection framework for Υ and i can be further improved by treating the regularization parameter as an additional selection parameter, as we will demonstrate in Section 7.

7 REFINEMENT: THE ROLE OF THE REGULARIZATION PARAMETER

As mentioned in Section 2.2 the regularization parameter α is important to control the smoothness of the orbital weights. However, up until now we have simply chosen α such that $\chi^2(\alpha)$ has roughly converged. In the following we will denote such a regularization

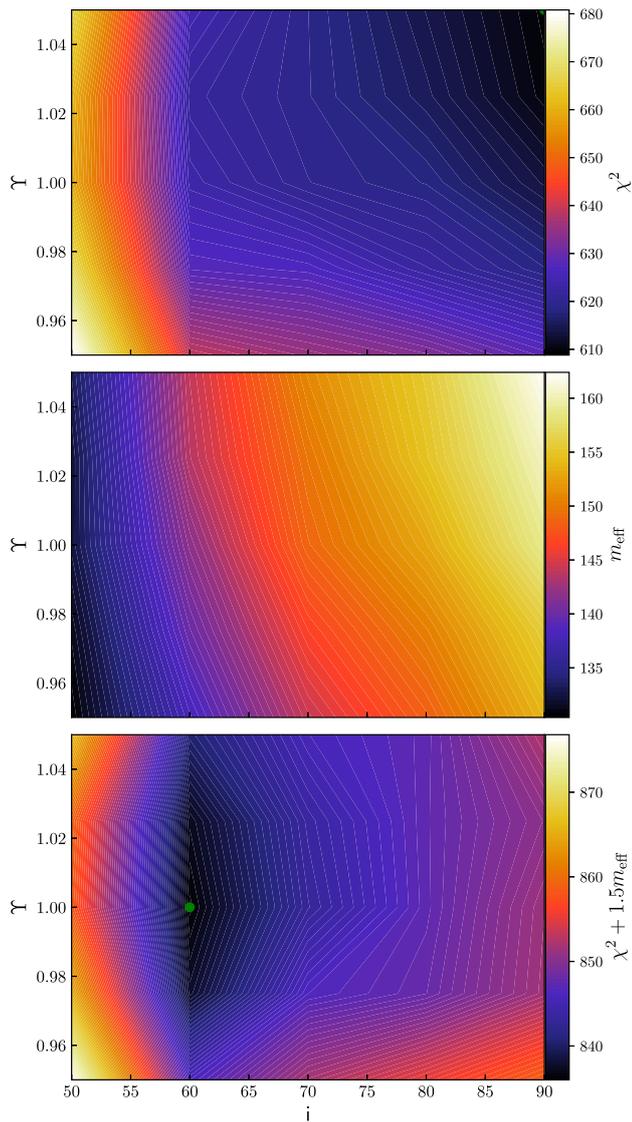


Figure 10. Top panel: The χ^2 distribution averaged over dynamical fits to 10 mock data sets of the flattened toy galaxy E with $i_{\text{true}} = 60^\circ$ and $\Upsilon_{\text{true}} = 1.0$. Middle panel: The corresponding distribution of the average number of effective parameters. Bottom panel: The model selection distribution with the calibrated approach (equation 14 with $w_m = 1.5$). The green dot locates the parameters of the best model. A simple χ^2 minimization leads to a bias in i and Υ because models with larger i and larger Υ are more flexible. Model selection removes this bias and allows to clearly identify the correct model.

parameter that was chosen using this convergence criterion as α_∞ . For the modelling of the toy galaxies in Table 1 this happens for $\alpha \gtrsim 1$, which motivated our somewhat arbitrary choice of $\alpha_\infty = 1.67$. This means we almost only considered the most flexible models as candidate models, since $\chi^2(\alpha)$ has roughly converged at this regularization value, meaning a further increase in α does not significantly reduce χ^2 further and the corresponding number of effective parameters has plateaued (cf. Fig. 3). While this choice guarantees a good fit to a given kinematic data set, it may not be suitable to restrict our test only to models with such a large α as they are prone to overfitting the data. And this, in turn, could potentially weaken the constraints on the selection parameters. Moreover, the optimal weight w_m of the m_{eff} in the model evaluation

(cf. equation 14) could depend on α , meaning that a calibration found for one data set of one galaxy may not be applicable to another data set of another galaxy. It is common practice to use Monte Carlo simulations of the galaxy under investigation to determine its optimal smoothing (e.g. Saglia et al. 2000; Cretton & Emsellem 2004; Thomas et al. 2005; Morganti & Gerhard 2012; Neureiter et al. 2021). In order to avoid smoothing-induced biases in the models, the mock data set should be as realistic as possible and adapted to the particular data set at hand (resolution, signal-to-noise, etc.). In addition, since the optimum value of α will depend on the underlying galaxy structure such simulations need to be repeated for every galaxy anew. However, a more targeted approach that optimizes α directly from the observed data may be advantageous as it would not rely on the choice of mock galaxy that is required for the Monte Carlo simulations.

For this purpose we again employed the concept of selection parameters. Section 2.1 characterized selection parameters as those that need to be specified to single out an individual model out of the family of models $\mathcal{F}(M_\bullet, \Upsilon, i, \dots)$. The respective orbital weights w_i are its free parameters. In general, there may exist many distribution functions (or equivalently vectors \mathbf{w}) that satisfy the observational constraints posed by $\chi^2(\mathbf{w})$ for a given set of selection parameters. The reason is that $\chi^2(\mathbf{w})$ is not necessarily strictly convex (cf. Neureiter et al. 2021). In practice, this issue is circumvented by adopting a penalized maximum-likelihood estimation by adding a penalty function, such as the entropy term in equation (5) that makes the solution for the orbital weights unique. However, this means out of all the allowed and statistically viable DFs only a single, privileged DF is identified, a possible issue already pointed out by Magorrian (2006). In the following we will show that our model selection framework facilitates an evaluation of multiple distribution functions compatible with a single set of $M_\bullet, \Upsilon, i, \dots$ if we extend the selection parameters to include the regularization parameter α .

Which distribution function out of all the compatible DFs (or \mathbf{w}) is privileged by a penalized maximum-likelihood approach is determined by the specific value of the weight parameter used for the penalty function (in our case α). In that sense the regularization parameter α is a prior that, like the selection parameters, constrains the parameter space available to the orbital weights. Furthermore, like models with different selection parameters, models with different α also have varying number of effective parameters. In fact, α plays the dominant role in determining the flexibility of a dynamical model, as demonstrated in Section 3.1. Consequently, we may attempt to extend the concept of selection parameters to include prior parameters, such as α , that push the orbital weights towards certain solutions. The regularization does not necessarily have to be the only such prior parameter that is used to identify a specific DF out all the DFs compatible with a set of selection parameters ($M_\bullet, \Upsilon, i, \dots$). For example, it is not necessary to use the phase volumes V_i of the orbits within the entropy-term. Instead one could use a free ω_i to bias the solution towards a specific solution (cf. Neureiter et al. 2021). In that case the ω_i are simply another set of selection parameters that could potentially be constrained by the establishment of candidate models with different ω_i . However, in our case the phase volumes are not themselves independent selection parameters as they are completely determined by the other selection parameters ($M_\bullet, \Upsilon, i, \dots$) that were used to create an orbit library. Therefore we focus on the regularization parameter α that is used to control the smoothness via the entropy term in equation (5).

If α can be thought of as simply another selection parameter, it may be possible to constrain it using the kinematic data by

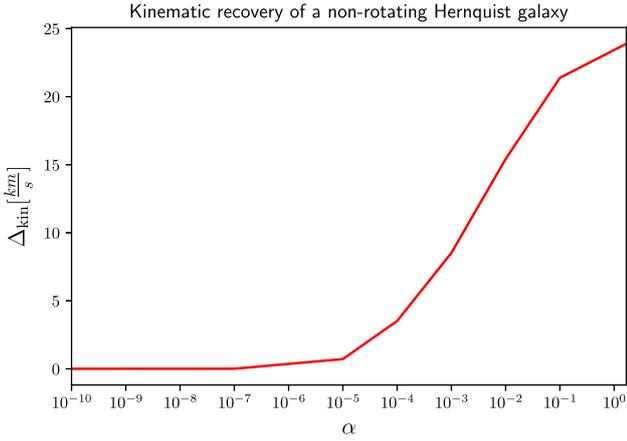


Figure 11. The total deviation in the intrinsic first and second velocity moments (equation 17) of fits to the non-rotating Hernquist galaxy (Galaxy A) for different values of the regularization parameter α . The inclination and mass-to-light ratio were fixed at their true values ($\Upsilon = 1.0$, $i = 60^\circ$). Larger α imply less influence of the entropy term and lead to an overfit of the noise. This increases the discrepancy between input model and fit. The lowest α allow an almost perfect reconstruction of the input model.

constructing candidate models with various α , analogously to any other mass or library parameter. While a recovery using only the kinematic goodness-of-fit is not sensible as it would always favour the maximum α models, it may be possible to constrain the regularization by evaluating the models within a model selection framework that takes into account the number of effective parameters. To investigate this possibility we modelled 10 mock data sets of toy galaxy A again, but this time we varied not only the two selection parameters Υ and i , yet also the regularization α . Then, as described in Section 5, for each model defined by (α, Υ, i) we estimated m_{eff} and selected the best model via the general model selection approach $\chi^2 + w_m m_{\text{eff}}$.

Considering that Galaxy A was created using a maximum-entropy Schwarzschild model with $\alpha = 10^{-10}$ we would expect that an optimal model selection approach should return a model with the same $\alpha = 10^{-10}$ as the best model. The reason is that the orbital weights are uniquely determined if (α, Υ, i) are given. In general, models obtained with the same Υ and i but different α will have different orbital weights and so, by construction, $\alpha = 10^{-10}$ should return the weights closest to the input model. The regularization parameter itself is not a physically relevant parameter but simply a choice of prior that restricts the freedom of the orbital weights. However, as a consequence it significantly impacts the form of the resulting distribution function of a dynamical model and its respective kinematic moments. Therefore one should ideally use a regularization parameter that allows the best possible approximation of the true, underlying kinematic structure. To quantify this we characterized the difference between the underlying true orbit model and a fitted model by the average root-mean-square deviations (RMSDs) of the first and second-order velocity moments. Apart from the deviations of the individual moments we also define a total kinematic deviation:

$$\Delta_{\text{kin}} = \text{RMSD}_v + \text{RMSD}_{\sigma_r} + \text{RMSD}_{\sigma_t} \quad (17)$$

Here v is the mean stellar rotation velocity of the model around its z -axis, while σ_r and σ_t are the radial and tangential velocity dispersions.

Fig. 11 shows Δ_{kin} versus α when the inclination i and mass-to-light ratio Υ are fixed at the correct values. Obviously, the input

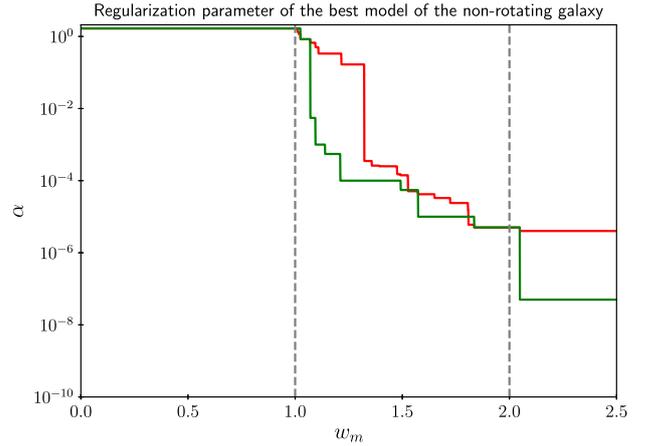


Figure 12. The mean (red) and median (green) regularization parameter of the best models to a non-rotating Hernquist mock galaxy (Galaxy A) as a function of the model-selection calibration parameter w_m . For every given w_m the α of the best model selected from all the possible candidate models with free selection parameters α , Υ and i is shown. The dashed vertical lines locate the model selection using the intuitive ($w_m = 1$) and the AIC ($w_m = 2$) approach (cf. equation 12 and 13). The case $w_m = 0$ corresponds to a simple χ^2 minimization and returns models with large α as the best ones. When the number of effective parameters is weighted by $w_m = 2$ (AIC), then the models identified as the best ones have much smaller α . For each of the 10 mocks AIC only selected models with $\alpha \leq 10^{-5}$. Such an α implies a very close match to the input model (Fig. 11).

model is exactly recovered for the lowest α . Larger α values lead to a larger discrepancy between the fitted velocity moments and the ones in the input model. The reason is that with increasing α the model starts to overfit the noise incorporated in the mock data. In analogy to Section 5, where we used the accuracy of the inclination recovery as criterion to decide upon the optimal weight factor w_m in $\chi^2 + w_m m_{\text{eff}}$ we can now also include α in the optimization of w_m . Fig. 12 shows the median and mean α values of the selected models for each mock of Galaxy A as function of the calibration weight w_m .

Unsurprisingly the simple χ^2 minimization selects the model with largest sampled α , in this case 1.67, meaning the best models overfit the noise and the recovered distribution function has a higher degree of irregularity than the smooth input model. For larger w_m we approximate the input model better as the median α shifts from 1.67 to 10^{-7} for $1 \lesssim w_m \lesssim 2$. In the regime of the Akaike approach, the selection framework yields $10^{-10} \lesssim \alpha \lesssim 10^{-5}$. In terms of the quality of the reconstruction of the internal velocity moments, any α in this regime leads to an almost perfect recovery of the input model (Fig. 11).

While the model selection framework with a variable α seems to work very well for the mocks of Galaxy A, in particular the Akaike approach, it may not be the most realistic set-up. For one because the generating model for Galaxy A is a maximum entropy model without any rotation, but mainly because one can generally not expect to have the ‘true’ generating model among the tested models, which in the case of Galaxy A is the model with selection parameters $i = 60^\circ$, $\Upsilon = 1.0$, $\alpha = 10^{-10}$. Furthermore, the generating $\alpha = 10^{-10}$ model of Galaxy A is also the model with the smallest number of effective parameters and consequently the selected model is always the generating model one as long w_m is chosen big enough, meaning that we only establish a lower boundary for the calibration weight w_m .

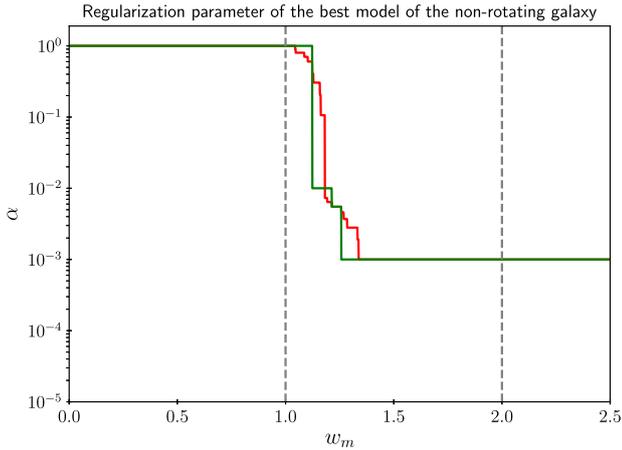


Figure 13. As Fig. 12, but for the rotating toy galaxy D. In contrast to the non-rotating galaxy, the model selection framework avoids maximum entropy models ($\alpha = 0$) here because such models would not allow an accurate reconstruction of the galaxy’s net rotation induced by its angular momentum bias $\lambda = 0.5$ (cf. Fig. 14).

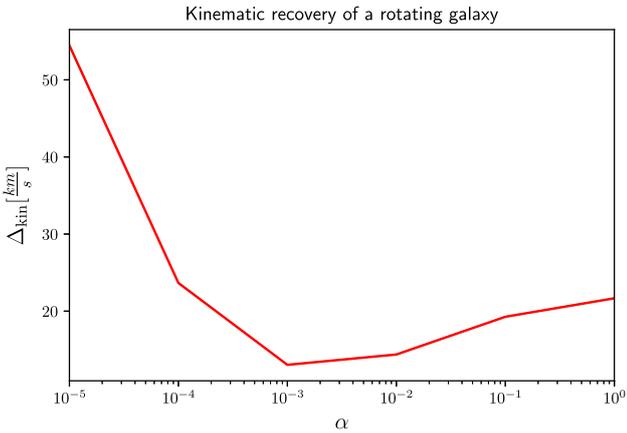


Figure 14. As Fig. 11, but for the rotating mock galaxy D. The recovery of the intrinsic kinematic moments is best at $\alpha = 10^{-3}$. This is exactly the value that the model selection framework identifies as best in the regime of the Akaike approach (Fig. 13).

For these reasons we additionally tested the approach with a variable regularization on the mocks of Galaxy D, which is a rotating toy galaxy with an angular momentum bias of $\lambda = 0.5$. In this case the best regularization parameter is not known in advance and the generating model is not necessarily among the tested models.

Fig. 13 shows the median and mean regularization parameter of the best model of Galaxy D as a function of the calibration weight w_m . We find large α for $w_m \lesssim 1$ followed by a steep falloff that turns into a stable plateau for $w_m \sim 2$, similar to Galaxy A. However, the optimal α for the Akaike approach is now at an intermediate $\alpha = 10^{-3}$ and not where the models have the smallest number of effective parameters. (Only for $w_m \gtrsim 4$ the approach tends towards models that simply have the smallest number of effective parameters. However, such large w_m can easily be discarded because they lead to unreasonably large $\chi^2 \gg N_{\text{data}}$, i.e. they are unable to fit the mock observations of Galaxy D).

The model selection framework $\chi^2 + w_m m_{\text{eff}}$ with $w_m \in [1.5, 4]$ suggests that the optimal amount of regularization for Galaxy D is achieved with $\alpha = 10^{-3}$. It becomes evident from Fig. 14 that

models with this regularization of $\alpha = 10^{-3}$ are indeed the models that reproduce the intrinsic velocity structure of Galaxy D best. This demonstrates that the model selection framework $\chi^2 + w_m m_{\text{eff}}$ with $w_m \in [1.5, 4]$ leads to an optimal recovery of the velocity structure of Galaxy D. Less regularized models, while nominally achieving a better χ^2 , overfit the LOSVDs and do not improve the recovered velocity moments. On the other hand more regularized models with $\alpha < 10^{-3}$ are not sufficiently flexible to be able to describe the underlying non-maximum entropy distribution function.

Treating the regularization as an additional variable selection parameter within our model selection framework, instead of arbitrarily fixing it to some value, does not merely affect the recovery of the intrinsic kinematics but also improves the constraining power of the other involved selection parameters. This is illustrated in Figs 15 and 16, which show the average, recovered mass-to-light ratio and inclination of Galaxy D for different model selection frameworks for two cases: (i) the often adopted approach where the smoothing parameter is fixed to some value (Fig. 15, where α is fixed to 1) and (ii) the more general case where α is a variable model parameter (Fig. 16).

As established in Section 6, a simple χ^2 minimization results in an overestimation of the mass-to-light ratio of about 3–4 per cent and an edge-on viewing angle for both, the modelling with fixed and the one with variable α . Only when the varying number of effective parameters of candidate models are accounted for with the model selection framework $\chi^2 + w_m m_{\text{eff}}$, one can find the correct selection parameters Υ and i . The constraining power of the model selection framework significantly improves if the regularization is treated as a selection parameter. In addition, the results are more stable with respect to the calibration weight w_m . In fact, compared to the case with fixed regularization, the Akaike approach yields unbiased results when the amount of regularization is optimized during the fit. This is further illustrated in the top panels of Fig. 17, where we show the average recovered selection parameters i and Υ against the calibration weight w_m for Galaxy A (left) and Galaxy D (right). With optimized regularization, the model selection results are stable beyond $w_m \gtrsim 1.5$, while modelling runs where we kept the regularization to a fixed value developed a bias to less flexible models beyond $w_m > 1.5$. For the modelling with variable α in the Akaike regime we further note a significantly reduced scatter of both Υ and i that appears to become smaller than our sampling steps.

Given the strong improvements on the constraints for the global selection parameters, we now investigate the intrinsic velocity distributions (the orbital anisotropy) under optimized regularization. In Fig. 18, we show the total kinematic RMSDs Δ_{kin} of the best models to Galaxy A and D with respect to the kinematics of the generating model. The figure also includes the RMSDs of the net velocity v and the anisotropy parameter $\beta = 1 - \sigma_t^2/\sigma_r^2$.

For both galaxies the approach with fixed and variable regularization are initially congruent for small $w_m \lesssim 0.3$ as the model selection is dominated by χ^2 and consequently, even in the case of variable regularization, models with α values as large as the one that we assumed for the fixed regularization case are favoured. When α is fixed, an increase of w_m beyond $w_m \gtrsim 0.3$ counteracts the inclination and mass-to-light ratio biases, while the recovery of the intrinsic kinematics does not improve. For $\alpha \sim 1$ a ‘sweet-spot’ $1.2 \lesssim w_m \lesssim 1.5$ appears, where the correct inclination and mass-to-light ratio are obtained on average. This is in line with our calibration of the optimal w_m in $\chi^2 + w_m m_{\text{eff}}$ for $\alpha = 1.67$ described in Section 5. A further increase of w_m overcompensates the mass-to-light ratio and inclination biases, suggesting that the choice of the specific model selection framework would be crucial when modelling real galaxies.

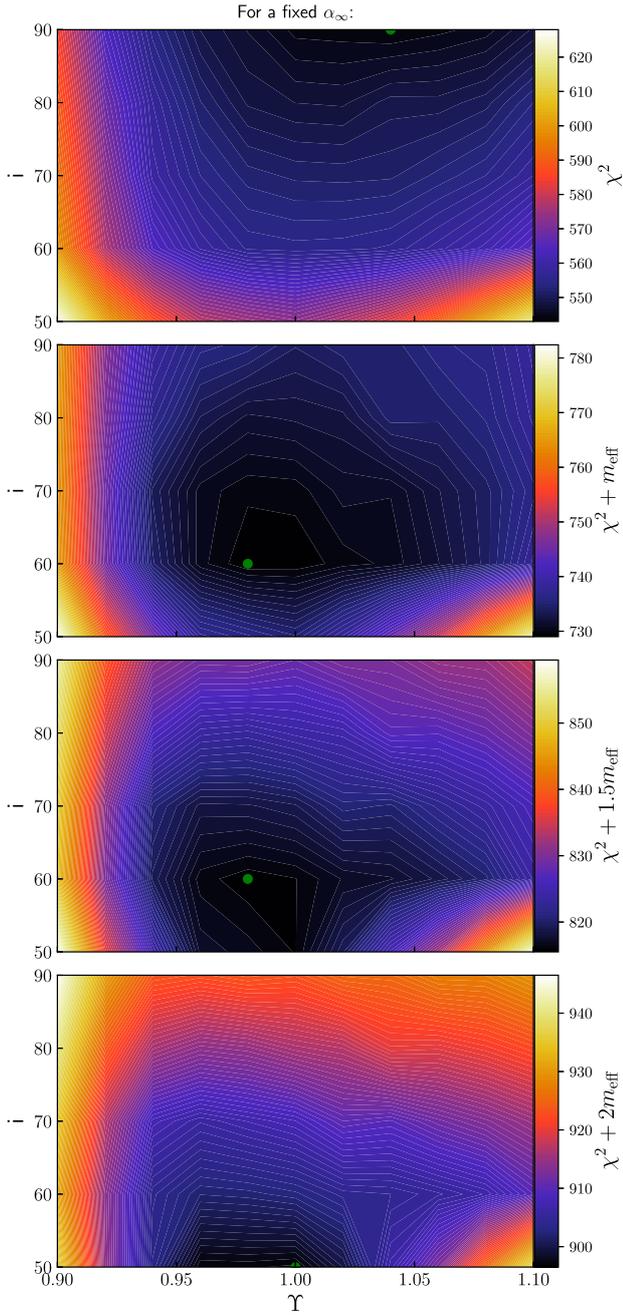


Figure 15. Constraints on the selection parameters i and Υ of Galaxy D for different model selection frameworks. From top to bottom: Simple χ^2 minimization, the ‘intuitive’ approach $\chi^2 + m_{\text{eff}}$, the calibrated approach $\chi^2 + w_m m_{\text{eff}}$ where w_m is determined from simulations (cf. Section 5) and the Akaike approach. In all cases, the smoothing parameter was held fixed to α_∞ , such that $\chi^2(\alpha)$ has approximately converged to a constant value. The green dot locates the best model in the respective model selection framework. The true values for Galaxy D are $i = 60^\circ$ and $\Upsilon = 1.0$. As expected, the χ^2 minimization leads to a biased result, the calibrated approach is best (because the simulations to calibrate w_m were made under the same assumptions for α). When α is fixed, the Akaike approach also leads to biases in the selection parameters, though these biases are smaller than the ones resulting from the simple χ^2 minimization.

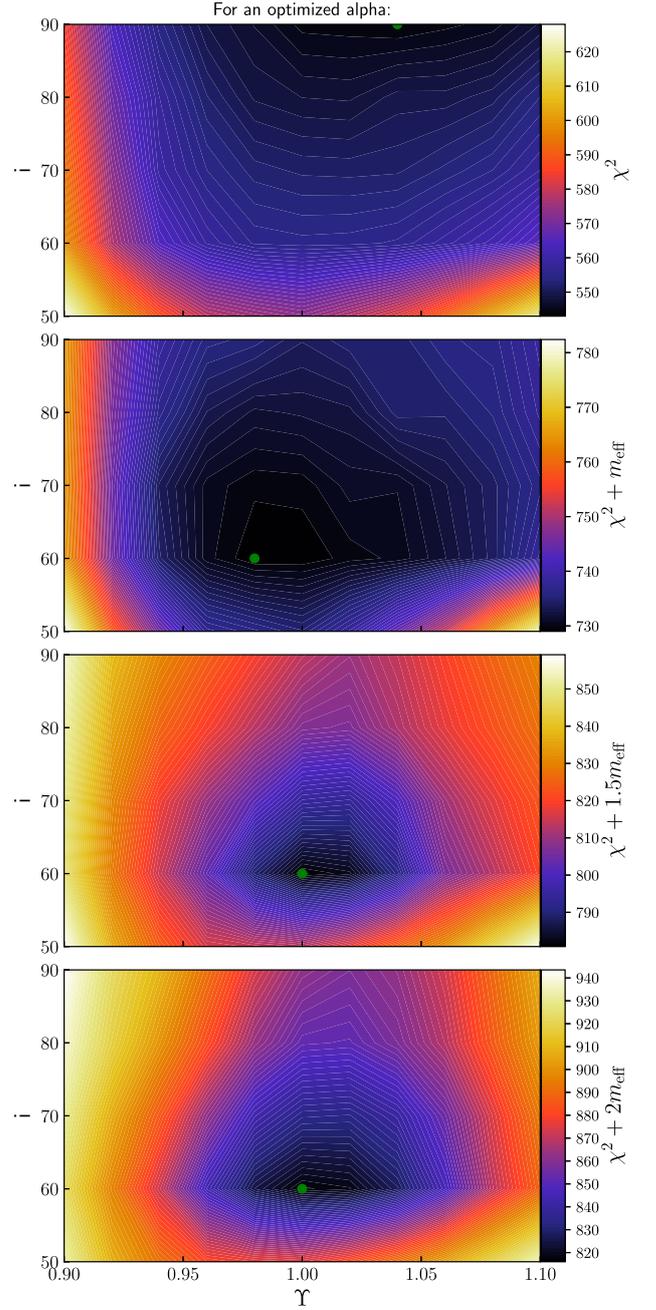


Figure 16. As Fig. 15, but the constraints on the selection parameters i and Υ of Galaxy D are shown for the case where α is treated as a free selection parameter. Compared to the case of a fixed α (Fig. 15) two important differences can be observed: (i) the constraints on the selection parameters have *improved significantly* and (ii) the results are stable, irrespective of the exact value of the weighting factor w_m in the model selection. In particular, the Akaike approach leads to an unbiased model selection now.

Remarkably, the choice of the calibration weight is simplified when the candidate models are optimized with respect to the degree of regularization: the correct selection parameters are recovered for a wide range of w_m , including the Akaike $w_m = 2.0$. Moreover, the recovery of the intrinsic kinematics, i.e. the rotation and the anisotropy in the second-order velocity moments, of both toy galaxies have *significantly improved* by treating the regularization as an additional selection parameter in the construction of the candidate

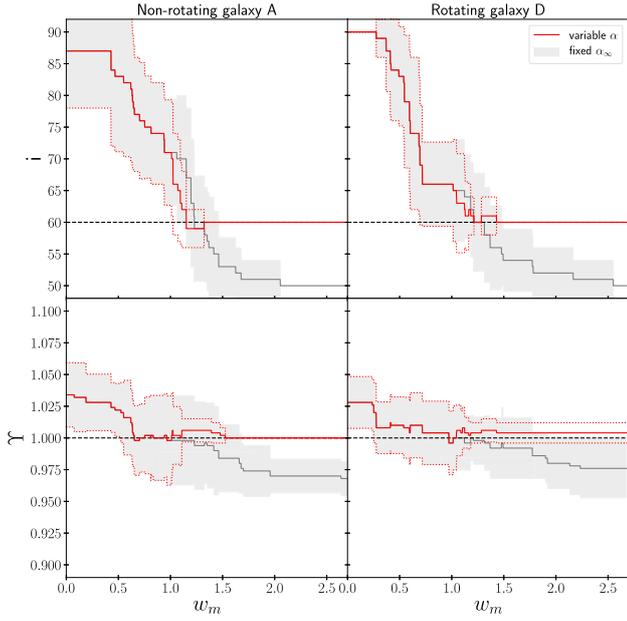


Figure 17. The recovered selection parameters of Galaxy A (left) and Galaxy D (right) using different model selection frameworks $\chi^2 + w_m m_{\text{eff}}$. Dashed, black line: The correct values for i and Υ . Solid, red Line: Recovered Υ and i averaged over the 10 mocks of the galaxy if the regularization parameter α is optimized like the other selection parameters during the fit. Dotted, red line: The corresponding $\pm 1\sigma$ error intervals. Solid, grey line: Recovered properties when modelling with a fixed regularization α_∞ . This α would be used if one adopts the $\chi^2(\alpha) \rightarrow \text{const.}$ criterion to choose the amount of regularization. Grey regions: The corresponding $\pm 1\sigma$ confidence regions. For models with fixed regularization, there is a relatively small range of w_m which lead to completely unbiased results for both, Galaxy A and D. Specifically, the Akaike approach is not the optimal choice for weighting the m_{eff} . When the regularization is optimized, however, the model selection results become unbiased for a wide range of w_m above ≈ 1.5 and recover the correct selection parameters for both mock galaxies. Only when $w_m \gtrsim 4$ the results become biased again. This suggests that model selection is optimal in the Akaike regime as long as the regularization is optimized as well.

models. With the right model selection, the intrinsic anisotropy can be recovered more accurately with uncertainties no larger than $\Delta\beta \sim 0.04$ in the mean.

Simply fixing the regularization complicates rather than simplifies the selection of the best candidate model and in the worst case may even skew the results for other selection parameters. Therefore dynamical modelling of galaxies with an a priori fixed amount of regularization should be avoided if possible. For example, Thomas et al. (2007b) show for their sample of Coma galaxies that whether a galaxy is radially anisotropic or tangentially anisotropic is rather independent of the assumed regularization. However, the strength of the actual anisotropy, i.e. how much radial or how much tangential the orbit distribution is, does indeed depend on the amount of regularization applied in the models.

8 A REAL GALAXY: DYNAMICAL MODELLING OF NGC 3368

We now turn to the application of our model selection approach to the real galaxy NGC 3368. This galaxy is part of the SINFONI black hole survey (cf. Saglia et al. 2016) and was chosen by us as follow-up of a previous analysis by Nowak et al. (2010).

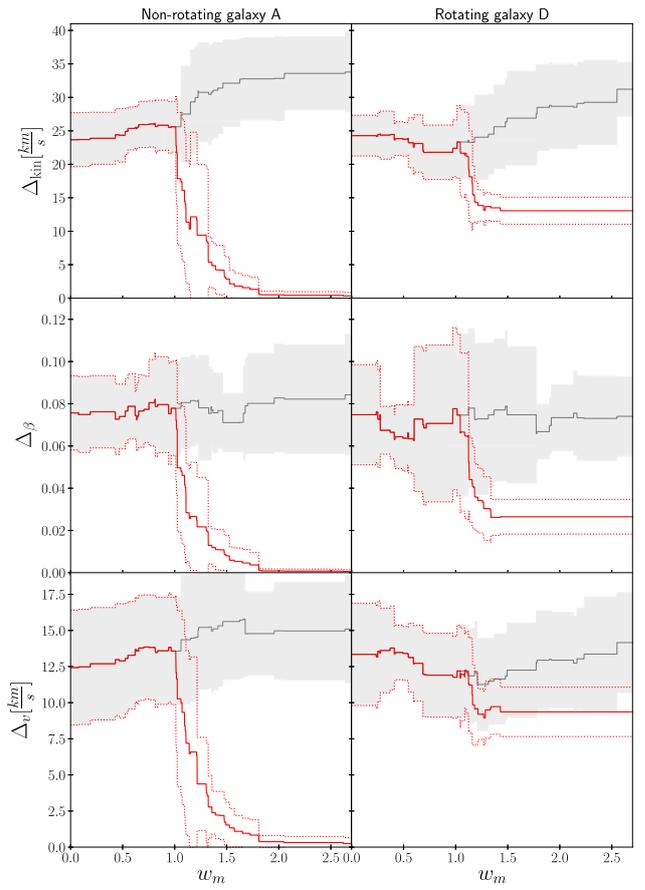


Figure 18. Similar to Fig. 17, but for the recovery of the intrinsic velocity moments of galaxies A (left) and D (right). Top panel: The total kinematic Root-mean-square deviations Δ_{kin} of the models selected with $\chi^2 + w_m m_{\text{eff}}$. Middle panel: The corresponding RMSDs Δ_β of the anisotropy parameter β . Bottom panel: The RMSDs for the net rotation, Δ_v . Both the anisotropy and rotation are emulated well only if α is optimized simultaneously to the other selection parameters and the best model is selected using $\chi^2 + w_m m_{\text{eff}}$ with a w_m in the Akaike regime. For Galaxy D the Akaike approach with α as a free selection parameter also selects models with intrinsic moments closer to the true kinematics, albeit not as close as it is the case for Galaxy A.

Due to the fact that NGC 3368 is a disc galaxy, its inclination can be estimated independently of the dynamical modelling. It is therefore a useful test bed for the inclination recovery. The observed ellipticity $\epsilon \approx 0.37$ of the outer disc implies a minimum inclination of $i_{\text{min}} = 51^\circ$ for an axisymmetric razor-thin disc. Assuming a typical (intrinsic) disc axial ratio of $q \approx 0.2$, NGC 3368 is expected to be inclined at a slightly larger angle of $i = 53^\circ$ to project to the observed ellipticity. Nowak et al. (2010) also estimated the inclination of NGC 3368 via the Tully–Fisher relation, which yields $i \approx 48^\circ$.

In Section 8.1 we will provide a short overview of the imaging and spectroscopy data. More details can be found in Nowak et al. (2010). Section 8.2 reviews the modelling results obtained from a traditional χ^2 minimization at fixed regularization. Model selection results with optimized regularization follow in Section 8.3.

8.1 Overview – photometric and kinematic data for NGC 3368

NGC 3368 is a double-barred spiral galaxy with a composite bulge consisting of a larger pseudo-bulge and a smaller classical bulge (cf. Erwin 2004). The galaxy is classified as a LINER2 (Ho, Filippenko &

Sargent 1997). The activity is so weak that we do not expect it to influence the dynamical modelling results. We assume that NGC 3368 lies at a distance of $d = 10.4$ Mpc (Tonry et al. 2001) where 1 arcsec corresponds to 50.4 pc.

The SINFONI kinematics of NGC 3368 were obtained with the K -band grating in the 100mas resolution mode. The average adaptive-optics corrected PSF had an FWHM of ≈ 0.165 arcsec. Non-parametric LOSVDs were obtained from the CO bandheads adopting the approach of Gebhardt et al. (2000). More details on the data reduction process, AO correction, selection of template stars, and the resulting kinematic maps of the galaxies can be found in Nowak et al. (2010).

The photometry acquired for NGC 3368 is a combination of different data sets: Sloan Digital Sky Survey observations (York et al. 2000) in the r -band were used to find the ellipticities of the outermost parts of the galaxy’s disc. Intermediate radii of NGC 3368 are mostly covered by a dust-corrected *HST* NICMOS2 F160W image and a K -band ($2.2 \mu\text{m}$) image of Knäpen et al. (2003) observed with the Isaac Newton Group Red Imaging Device (INGRID). The photometry is completed by the collapsed K -band data cube images taken with the integral field spectrograph SINFONI at the Very Large Telescope (VLT). For more detailed information on the image matching, seeing, dust-correction and an isophote analysis see Nowak et al. (2010).

As noted before, the photometric bulge of NGC 3368 has a composite structure. The central 2 arcsec are dominated by an almost round and kinematically hot ‘classical’ bulge. Outside this region a more flattened and kinematically cooler structure emerges: the discy pseudo-bulge. Since the SINFONI FoV has a size of 3 arcsec \times 3 arcsec, both components are relevant for the modelling. To account for potentially different mass-to-light ratios in the respective stellar populations, we deproject the classical bulge separately from the pseudo-bulge and assign each deprojection its own mass-to-light ratio Υ . To this end, we adopt the photometric bulge–disc decomposition of Nowak et al. (2010). It assumes that the discy pseudo-bulge is the inner extension of the galaxy’s large-scale disc and combines both to a single photometric disc component. The other photometric component, respectively, consists only of the ‘classical’ bulge. In practice, the deprojections were again obtained using the Metropolis-Algorithm of Magorrian (1999) without any shape prior. Non-axisymmetric features in the photometric data, like bars and spiral arms, are averaged over as described in Nowak et al. (2010). We tested a grid of assumed inclinations ranging from $i = 53^\circ$ to $i = 90^\circ$. In the following we simply refer to the classical bulge component as the bulge, and the pseudo-bulge + disc as the disc component.

8.2 χ^2 -modelling of NGC 3368

This section illustrates the results of the ‘traditional’ dynamical modelling approach for NGC 3368, which is based (i) on evaluating different mass models based on a pure χ^2 -analysis (i.e. assuming $m_{\text{eff}} = \text{const}$) and (ii) on a fixed strength of the regularization. These results are later contrasted to the model selection framework with optimized regularization (Section 8.3).

For our final mass model we add a central SMBH to the two stellar mass components:

$$\rho = M_\bullet \cdot \frac{\delta(r)}{4\pi r^2} + \Upsilon_{\text{bulge}} \cdot \nu_{\text{bulge}}(i) + \Upsilon_{\text{disc}} \cdot \nu_{\text{disc}}(i), \quad (18)$$

where $\nu_{\text{bulge}}(i)$ and $\nu_{\text{disc}}(i)$ are the luminosity densities at assumed inclination i . For the χ^2 analysis of this section we considered only two inclinations: $i = 53^\circ$ and $i = 90^\circ$. A denser inclination grid is used for the model selection approach (Section 8.3). Even though

it seems likely that a considerable amount of the total galaxy mass is contributed by a dark matter (DM) halo, the addition of a DM component to this model of the galaxy centre is not required. Erwin et al. (2018) demonstrated that the addition of a DM halo to a two-component stellar model of (the inner parts of) a disc galaxy neither improves the fit significantly, nor does it change the black hole mass M_\bullet or bulge mass-to-light ratio Υ_{bulge} drastically. It’s only that the disc mass-to-light ratio Υ_{disc} of such a two-component model without DM is larger than the Υ_{disc} of a mass model that does include DM (and probably larger than the actual Υ of the stars in the disc). This is because the disc ‘absorbs’ the dynamical role of the ignored dark mass, which leads to a Υ_{disc} that is biased high.

With the above mass model we constructed trial models by varying the three traditionally relevant selection parameters that determine the orbit library: the mass-to-light ratios of bulge and disc, and the black hole mass. With the goal to achieve an efficient, yet none the less sufficiently dense sampling of these parameters we narrowed down the approximate location of the χ^2 minimum with an initial round of trial models that sparsely sampled a wide range of mass parameters. Using the information gained from this initial χ^2 minimization we decided for the following final sampling grid: The black hole mass is sampled linearly in the interval $[1.0 \times 10^6 M_\odot, 19.0 \times 10^6 M_\odot]$ with a stepsize of $2.0 \times 10^6 M_\odot$ and the bulge mass-to-light ratio is linearly sampled 15 times in the interval $[0.20, 0.90]$. It turned out that the SINFONI FoV is too small (inner 3 arcsec of the galaxy) to constrain Υ_{disc} . Therefore we decided to sample only $\Upsilon_{\text{disc}} \in \{0.2, 0.4, 0.6\}$ to save computation time. This indeterminacy of Υ_{disc} was already noted in Nowak et al. (2010). For the smoothing we applied the old approach to set $\alpha = 1.67$ guaranteeing that $\chi^2(\alpha)$ has converged.

Estimating the *statistical* errors of the dynamical models is non-trivial, mostly because unlike in our toy galaxies, the measurement errors of real LOSVDs are strongly correlated (e.g. Houghton et al. 2006). If these correlations are unknown or neglected then the calculated χ^2 is systematically smaller than $N_{\text{data}} - m_{\text{eff}}$ (cf. Tables 2 and 3). The absolute χ^2 values and, consequently, also the $\chi^2 + 2 m_{\text{eff}}$ significance intervals become inaccurate then. Ideally this issue could be solved by taking the full covariance matrix of the observations into account in the modelling, we plan to investigate this in the future. Another way to circumvent the issue would be the modelling of multiple observations of the same galaxy. Then one could calculate the RMS of the recovered selection parameters, analogously to the modelling of 10 different mocks of our toy galaxies (e.g. Section 5). In practice, this is unfeasible. However, if the investigated galaxy is perfectly axisymmetric then each quadrant provides essentially an independent observation of the same underlying galaxy structure. This simple approach of error estimation can also be adopted for galaxies that deviate from axisymmetry, in which case the resulting RMS will be increased by the systematic structural differences between the quadrants (cf. Nowak et al. 2010; Rusli et al. 2013).

Fig. 19 illustrates the resulting χ^2 constraints for the mass-to-light ratios and black holes mass for both, the 53° models and the 90° models in the four galaxy quadrants. The models obtained under an assumed inclination of $i = 53^\circ$ (motivated by the disc flattening as described above) yield black hole masses $M_\bullet \in [3 \times 10^6 M_\odot, 5 \times 10^6 M_\odot]$. The formal average over the four quadrants is $M_\bullet = (4.0 \pm 1.0) \times 10^6 M_\odot$. However, in some quadrants black hole masses of up to $13 \times 10^6 M_\odot$ are hardly ruled out when looking at the detailed χ^2 curves. The mass-to-light ratio of the bulge is well constrained with an average value of $\Upsilon_{\text{bulge}} = 0.59 \pm 0.07$. As already mentioned, the disc mass-to-light ratio is essentially unconstrained by the SINFONI kinematics. The χ^2 constraints for the black hole

Table 2. Best-fitting models according to a simple χ^2 minimization. Columns 1 and 2: the modelled quadrant of NGC 3368 and the assumed inclination. Columns 3–5: K -band mass-to-light ratios Υ_{bulge} , Υ_{disc} , and black hole mass of the best-fitting model. Columns 6 and 7: the χ^2 and χ^2/N_{data} of this model. $N_{\text{data}} = 945$.

Quadrant	i	$\Upsilon_{\text{bulge}} (M_{\odot}/L_{\odot})$	$\Upsilon_{\text{disc}} (M_{\odot}/L_{\odot})$	$M_{\bullet} (10^6 M_{\odot})$	χ^2	$\frac{\chi^2}{N_{\text{data}}}$
1	53°	0.65	0.20	5.0	363.230	0.3844
2	53°	0.55	0.20	5.0	479.117	0.5070
3	53°	0.50	0.40	3.0	407.302	0.4310
4	53°	0.65	0.40	3.0	347.183	0.3674
1	90°	0.75	0.20	1.0	321.628	0.3403
2	90°	0.70	0.40	5.0	460.396	0.4872
3	90°	0.70	0.20	1.0	405.026	0.4286
4	90°	0.50	0.40	15.0	331.003	0.3503

Table 3. Selection parameters derived from the model selection approach with optimized regularization for each quadrant of NGC 3368. The number of effective parameters m_{eff} has been weighted with $w_m = 2.0$ as in the Akaike information criterion and in accordance with the simulations in Section 7.

Quadrant	α	i	$\Upsilon_{\text{bulge}} (M_{\odot}/L_{\odot})$	$M_{\bullet} (10^6 M_{\odot})$	$(\chi^2 + 2.0 m_{\text{eff}})_{\text{min}}$
1	10^{-2}	65°	0.7	3.0	861.02744
2	10^{-2}	60°	0.4	11.0	934.158201
3	10^{-2}	53°	0.6	3.0	882.207510
4	10^{-2}	53°	0.5	5.0	812.338033

mass at $i = 90^{\circ}$ are much noisier than at $i = 53^{\circ}$. Averaged over the four quadrants, we find $M_{\bullet} = (5.5 \pm 5.7) \times 10^6 M_{\odot}$ (with a more than five times larger scatter than in the $i = 53^{\circ}$ case). In fact, neither very small black hole masses $M_{\bullet} < 1 \times 10^6 M_{\odot}$ nor very large ones $> 15 \times 10^6 M_{\odot}$ can be ruled out at a significant confidence level. While we do not observe such a strongly increased scatter in Υ_{bulge} we do see evidence for a slight shift towards larger mass-to-light ratios when modelling the galaxy edge-on: $\Upsilon_{\text{bulge}} = 0.66 \pm 0.10$.

In the case of NGC 3368 the observed disc ellipticity strongly suggests that the models at $i = 53^{\circ}$ are more realistic than the ones at $i = 90^{\circ}$, but we do not have such prior information about a galaxy's inclination in every case, in particular not if it is a generic early-type galaxy. In that case one would be reliant on the χ^2 values of the differently inclined models. Table 2 shows the selection parameters and corresponding minimum χ^2 values of the best-fitting model for each quadrant and inclination. In all quadrants, the edge-on models have a lower χ^2 value than the models at $i = 53^{\circ}$. In other words, if NGC 3368 would be modelled without external knowledge about its inclination a pure χ^2 analysis would erroneously conclude that the galaxy is seen edge-on, when, in fact, an independent inclination measurement would indicate $i \approx 48^{\circ}$ – 55° . These results are not surprising. In Section 5 we have seen that a simple χ^2 minimization tends to yield results that are biased towards $i = 90^{\circ}$. Moreover, in Section 6 we have seen that this inclination bias comes along with a noticeable mass bias (of the order of 10 per cent). Assuming that NGC 3368 is seen under an inclination angle close to the expected $i = 53^{\circ}$, than our χ^2 analysis of NGC 3368 is fully consistent with our expectations based on the simulated toy galaxies discussed in the previous sections.

8.3 Model selection of NGC 3368

In Section 7 above, we have demonstrated that the most accurate reconstruction of a galaxy's mass distribution and internal structure is achieved when using a model selection approach rather than a simple χ^2 minimization. Moreover, treating the regularization strength in a similar way as other selection parameters turned out (i) to improve the

accuracy of the galaxy reconstruction and (ii) to reduce the sensitivity of the results on the weight w_m for the effective parameters m_{eff} . Here, we transfer this approach to NGC 3368 and minimize $\chi^2 + 2.0 m_{\text{eff}}$. The choice for the Akaike weight $w_m = 2.0$ directly follows from the results of Section 7. We adopt the same mass model (equation 18) as for the χ^2 analysis in the previous Section 8.2, consisting of a central black hole, a disc and a bulge. Neglecting potential effects of the deprojection degeneracy this implies that the orbit models are completely determined by the five selection parameters M_{\bullet} , Υ_{bulge} , Υ_{disc} , i , α . However, in order to reduce the computation time we restricted ourselves to models with a fixed $\Upsilon_{\text{disc}} = 0.4$, since Υ_{disc} is almost unconstrained over the SINFONI FoV (Fig. 19).

Fig. 20 shows that the optimal amount of regularization is well constrained by the data: within the Akaike model selection framework all four quadrants of NGC 3368 are best modelled with an intermediate regularization of $\alpha = 10^{-2}$. Fig. 21 shows the corresponding constraints for Υ_{bulge} , M_{\bullet} and i (all results are listed in Table 3). Averaging the modelling results of the four quadrants we find that the AIC estimation yields a black hole mass of $M_{\bullet} = (5.5 \pm 3.3) \times 10^6 M_{\odot}$, a mass-to-light ratio $\Upsilon_{\text{bulge}} = (0.55 \pm 0.11) M_{\odot}/L_{\odot}$, and an inclination angle $i = (57.8 \pm 5.1)^{\circ}$. Compared to the results of the simple χ^2 -minimization in Section 8.2 we immediately see that the inclination bias has *disappeared*. The model selection approach with optimized regularization yields an inclination that is consistent with the inclination angle derived independently from the disc flattening. This result was expected after the analysis of the toy galaxies above. Still, it is somewhat astonishing given how small the FoV of the SINFONI kinematic data actually is. However, this underlines the power of the model selection approach to extract all the information contained in the 945 measured LOSVD data points in each quadrant. Another difference to the simple χ^2 minimization is the fact that the model selection approach yields a 10–15 per cent smaller Υ_{bulge} . Again, this is fully consistent with the toy galaxy results, where the recovered masses were biased high by up to ~ 10 per cent when a simple χ^2 minimization was performed, while there was no mass bias in the model selection framework. The only significant difference to

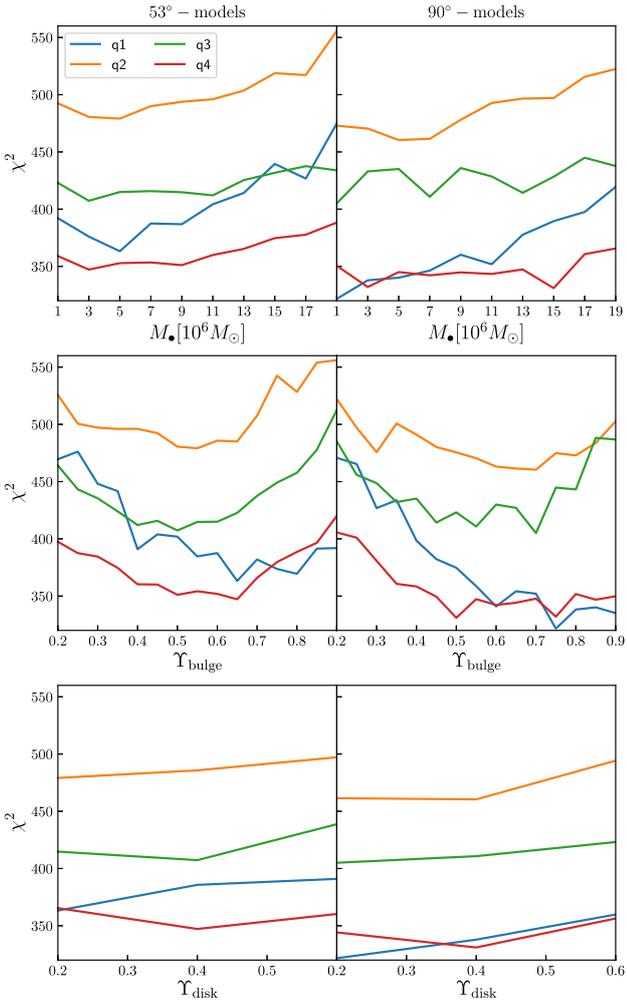


Figure 19. Left: χ^2 constraints for the mass parameters of the orbit models of NGC 3368 assuming an inclination of $i = 53^\circ$. Right: The corresponding χ^2 constraints assuming the galaxy is seen edge-on. We modelled the kinematic data of the four galaxy quadrants separately [coloured curves; the nomenclature of the quadrants is the same as in Nowak et al. (2010)]. For all four quadrants an edge-on model achieves a better fit than the more inclined model.

the simulation results is that the scatter in Υ_{bulge} of NGC 3368 is not reduced in the model selection framework compared to the χ^2 minimization. This could be caused by a multitude of issues:

- (i) Systematic differences between the quadrants of NGC 3368 reflecting the galaxy’s intrinsic non-axisymmetry.
- (ii) Inaccurate errors of the observed LOSVDs.
- (iii) Negligence of the correlations between the measurements of the same LOSVD at different line-of-sight velocities in the modelling and bootstrap iterations.
- (iv) A potential non-alignment of disc and bulge.

We used an updated version of our axisymmetric Schwarzschild code for this analysis, compared to Nowak et al. (2010). Therefore, the results of the pure χ^2 analysis cannot be compared easily. However, we note our black hole mass is slightly smaller than the $M_\bullet = (7.5 \pm 1.5) \times 10^6 M_\odot$ quoted in that paper while our Υ_{bulge} is about 25 per cent larger. These small discrepancies could also partly be caused by the different choice of mass parameter sampling and regularization.

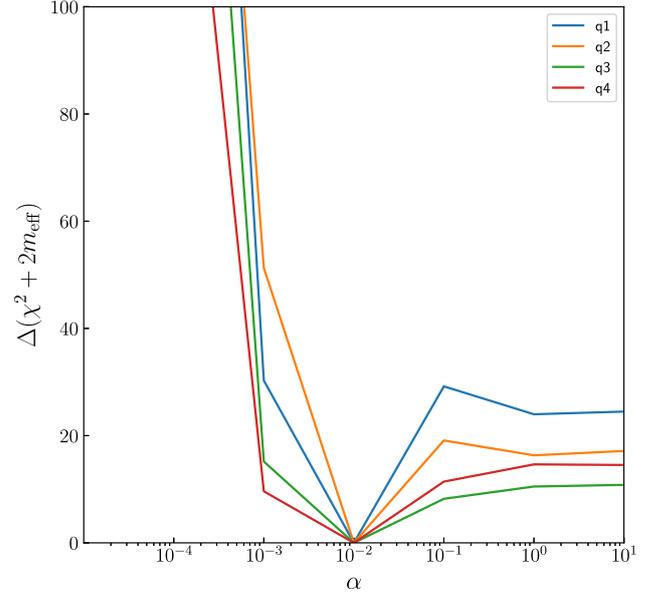


Figure 20. Constraints on the optimal smoothing for the orbit models of NGC 3368. At each α , we show the smallest achieved $\chi^2 + 2.0 m_{\text{eff}}$. Different colours represent models of different galaxy quadrants and the minimum $\chi^2 + 2.0 m_{\text{eff}}$ of each quadrant has been subtracted. The optimal smoothing is well determined in each quadrant.

After selecting a dynamical model with AIC one should always confirm that the selected model can in fact reproduce the observed data well as AIC only compares the models relative to one another but makes no statement about their absolute quality. Fig. 22 demonstrates this for the models of NGC 3368 by showing the major-axis kinematics of the best AIC and χ^2 models in relation to the observed SINFONI data. The error bars for the SINFONI data are the 1σ errors estimated from *uncorrelated* Monte Carlo realizations of the non-parametric LOSVD data, thus, they are slightly underestimated. Both, AIC and χ^2 models, fit the data well. However, as one would expect, the AIC model appears to be slightly smoother. It does not react to outliers as much as the χ^2 models do. Instead of a validation by eye as done here for the major axis of NGC 3368, it may be advantageous to cross-check the goodness-of-fit of AIC models more systematically by confirming whether $\chi^2 + m_{\text{eff}} \sim N_{\text{data}}$ holds for the selected models (cf. the discussion in the next Section 9). However, this is not a viable option as long as the correlations in the real LOSVDs are not implemented in the calculation of χ^2 and m_{eff} as this causes both values to be underestimated.

9 DISCUSSION

In the following we examine the *statistical* objective of AIC model selection in the context of dynamical modelling and address potential problems that may arise when estimating intrinsically degenerate properties with AIC (Section 9.1). Subsequently we review earlier studies attempting an inclination recovery using axisymmetric models (Section 9.2) and discuss the implications for the mass reconstruction when selecting models based on their χ^2 alone (Section 9.3).

9.1 Model selection and intrinsic degeneracies

The toy galaxy recoveries discussed in Sections 5 and 7 had in common that the selection parameters under investigation had a

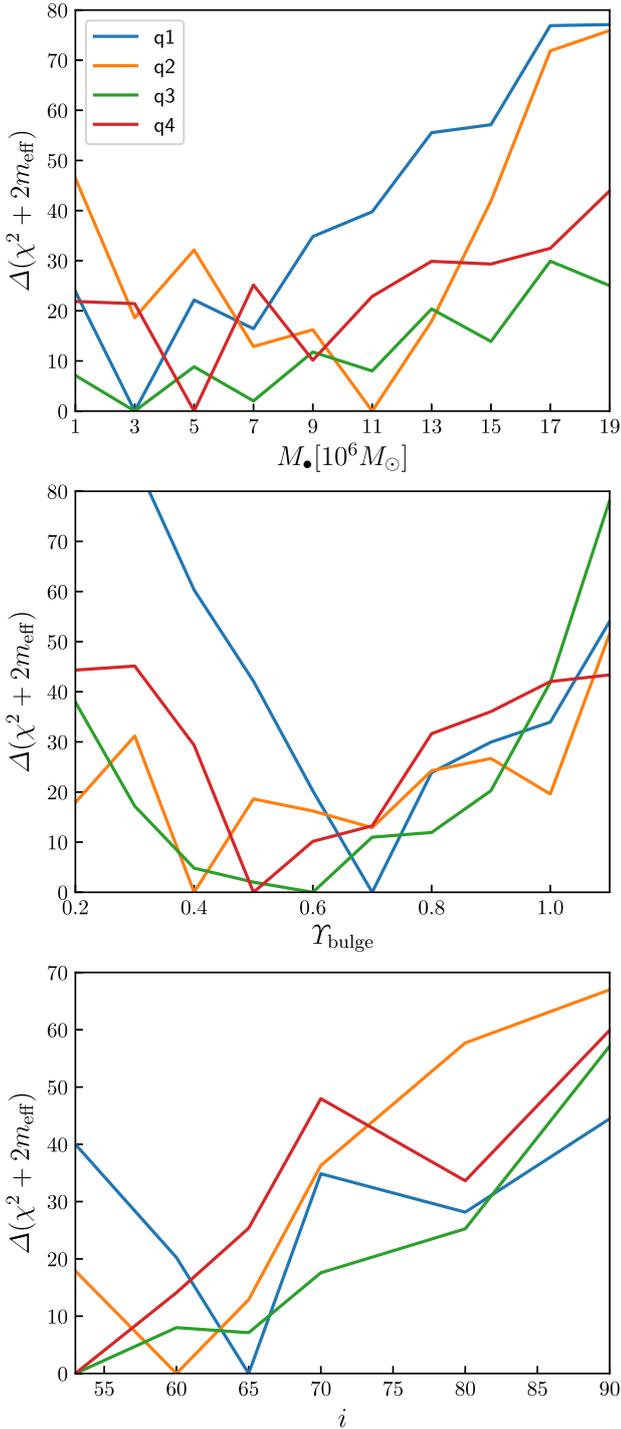


Figure 21. Similar to Fig. 20, but here we plot the black hole mass (top), bulge mass-to-light ratio (middle) and inclination i (bottom) against $\Delta(\chi^2 + 2.0m_{\text{eff}})$.

unique ‘true’ value. In the model selection set-up, this true value could be recovered in an unbiased way with high precision. In order to assess how the model selection framework operates in a degenerate case where such a true value does not exist, or where the data is insufficient to constrain a selection parameter, it is useful to get back to the example of the inclination recovery in a sphere (cf. Section 2.3): in a kinematically isotropic sphere all viewing angles should be equivalent. Fig. 23 shows $\chi^2 + m_{\text{eff}}$ and $\chi^2 + 2.0m_{\text{eff}}$ versus the

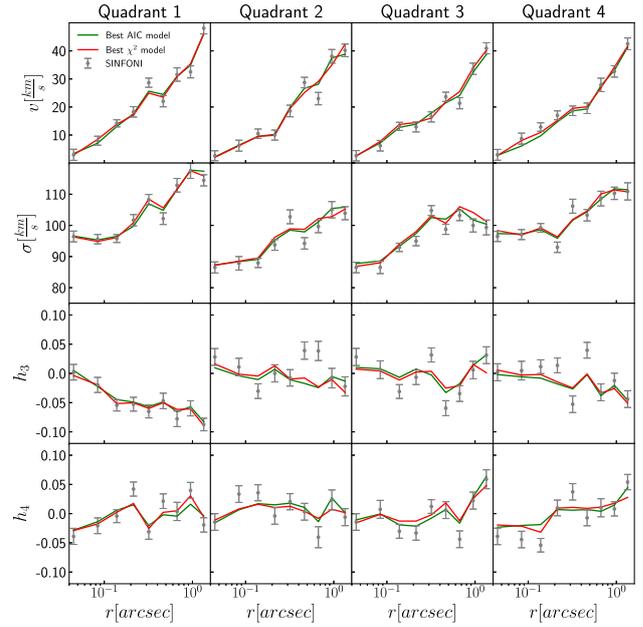


Figure 22. The Gauss-Hermite coefficients along the major axis of NGC 3368. Grey: SINFONI observations (the error-bars are estimated neglecting possible correlations in the LOSVD data; Red: Projected kinematics of the best χ^2 ; Green: Best model according to AIC.

inclination for the same N -body models as in Fig. 1 (cf. Section 2.3). At each tested inclination we optimized the regularization strength α as in Section 7. Apart from the regularization we only varied the inclination, meaning that the orbit libraries are the exact same for all models. In contrast to the non-spherical toy galaxies that we tested in the previous sections, the AIC approach now appears to be biased, in the sense that it prefers low inclinations although all viewing angles should be equivalent (lower panel of Fig. 23). The ‘intuitive’ model selection framework displays the degeneracy one would generally expect for a spherical model: no inclination is preferred over the other (upper panel of Fig. 23). We will come back to this below.

The reason why AIC assesses models of a spherical galaxy differently depending on the inclination can be understood by elaborating what the Akaike criterion is designed to achieve in a more general, statistical sense. Coming from information theory AIC estimates the information loss when modelling an underlying structure/process with a statistical model and selects the model that has the least information loss. The information loss of a statistical model is generally quantified by the Kullback-Leibler divergence (KLD) between the fitted model and the underlying (noise-free) structure it is supposed to represent. Naturally, since the latter is usually unknown, AIC is merely an estimate of the actual information loss.

When transferring these considerations to the context of dynamical models it is crucial to clarify what the *statistical* model and the underlying structure actually are in this case. The data being fitted by dynamical models are the observed kinematic data, implying that statistically speaking the underlying structure are the noise-free LOSVDs \mathbf{l}_{true} and not the galaxy’s distribution function or mass structure. Analogously, the statistical model is not the orbit model itself but is given by the LOSVDs \mathbf{l}_{mod} it produces (cf. Section 2). The internal structure of the model, i.e. how one arrives at \mathbf{l}_{mod} , is only of secondary concern when evaluating the models.

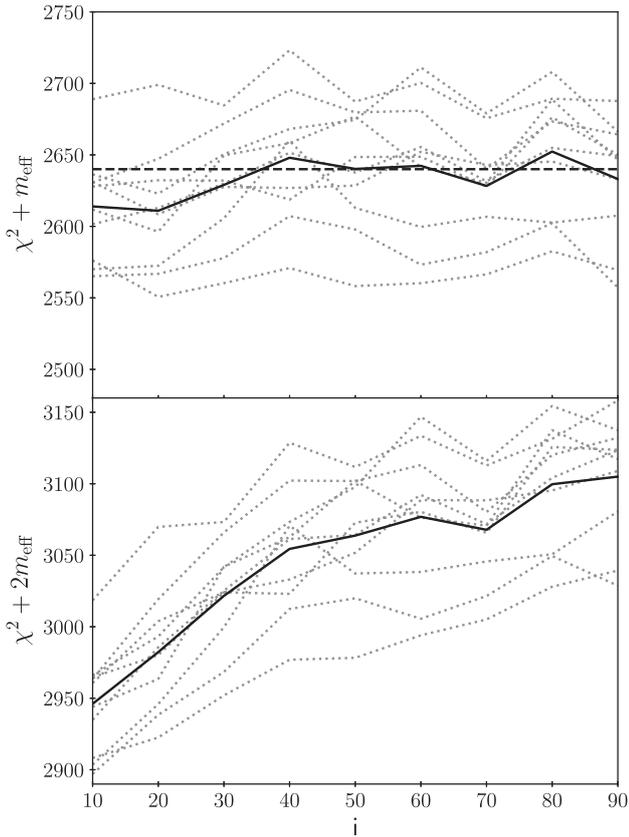


Figure 23. Model selection results for the Hernquist sphere (cf. Fig. 1 for the selection using only χ^2). Dotted lines: The selected best models for the individual mocks. Solid Line: The corresponding mean. Dashed line: The number of kinematic data points. For this spherical galaxy AIC appears to favour the models with the smallest inclination which also have the smallest number of effective parameters. The $\chi^2 + m_{\text{eff}}$ curves exhibit the inclination degeneracy one would expect for a spherical galaxy: $\chi^2 + m_{\text{eff}} \sim N_{\text{data}}$.

What this means for the modelling of the isotropic N -body sphere is that the AIC is supposed to minimize the differences between the model LOSVDs and the true LOSVDs. As mentioned above, this difference is quantified by the KLD from I_{mod} to I_{true} but can also be illustrated more heuristically by the RMSD ΔI of the LOSVDs.

As shown in the top panel of Fig. 24, ΔI behaves indeed very similar to $\chi^2 + 2m_{\text{eff}}$ (cf. Fig. 23): both have their minimum where the orbit library is viewed close to face-on. In that sense the AIC selection achieves exactly what it is intended to do, namely to select the model that emulates the underlying LOSVDs best. That the true LOSVDs happen to be represented slightly better at one inclination than another is related to the design of our model. While the sphere itself has no preferred viewing angles, our axisymmetric model *does* have such a preferred axis. For example, while we have pairs of orbits that only differ in the sign of L_z we do not have the equivalent pairs of orbits that only differ in the sign of other angular-momentum components. We expect that an orbit sampling using a 5d starting space (Neureiter et al. 2021) will lead to a different behaviour with respect to the assumed viewing angles in a sphere.

Irrespective of the question which particular inclination happens to yield the *best* LOSVDs we can ask whether a selection parameter (here: the inclination) is constrained by the data or not. As already mentioned above, the top panel of Fig. 23 shows that $\chi^2 + m_{\text{eff}} \sim N_{\text{data}}$ at all inclinations. This is exactly the expected behaviour when

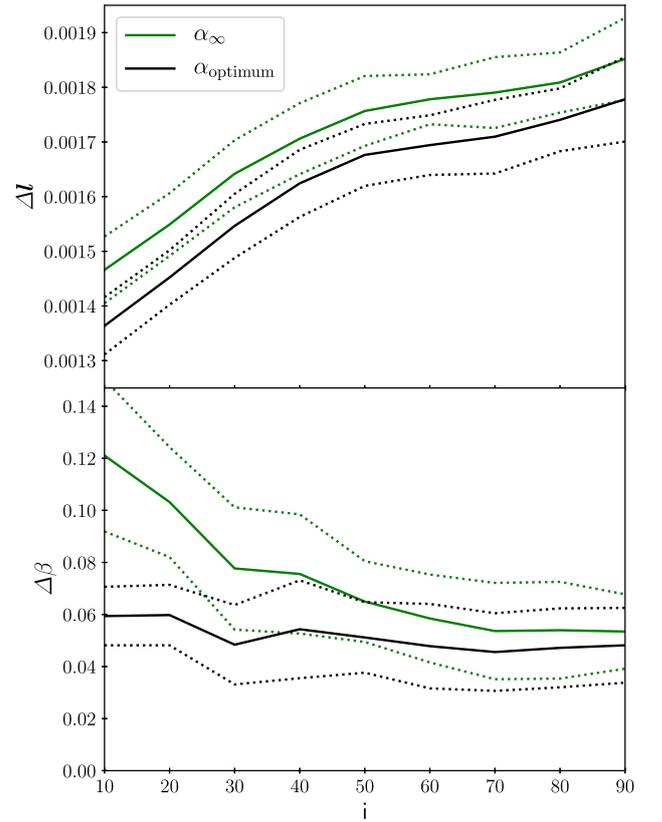


Figure 24. Top panel: The LOSVD RMSDs of the orbit models for the Hernquist sphere. Solid lines: The arithmetic mean of the 10 mocks. Dotted line: The corresponding estimated 1σ errors. Models with α_∞ are shown in green and models where the regularization was optimized using AIC are shown in black. These latter models are the same as in Fig. 23. Bottom panel: As the upper panel but for the RMSD of the anisotropy parameter β . Selecting models by their AIC values minimizes the information loss and leads to an optimized recovery of the underlying LOSVDs, unlike a selection based on χ^2 alone which prefers edge-on models with α_∞ . For the spherical galaxy the models that approximate the Hernquist LOSVDs best are at $i = 10^\circ$ and have $\alpha_{\text{optimum}} = 10^{-2}$. However, at fixed regularization, improving the recovery of the LOSVDs does not necessarily entail an improved recovery of the internal kinematics as illustrated by the green lines.

the data does not constrain the exact value of i : For a viable statistical model we expect $E(\chi^2) = N_{\text{data}} - m_{\text{eff}}$ (Section 3) and if this can be achieved for all i , then the data does not constrain the inclination. Among all these viable models the model selection picks up the one with the smallest m_{eff} . This is a generic result, because when $\chi^2 + m_{\text{eff}} \sim N_{\text{data}}$, then

$$\chi^2 + 2m_{\text{eff}} \sim N_{\text{data}} + m_{\text{eff}}. \quad (19)$$

In other words, when a selection parameter is not constrained by the data and $\chi^2 + m_{\text{eff}} \sim N_{\text{data}}$ over some extended interval, then the model selection may be biased in the sense that it will pick up that particular model inside the interval which has the lowest m_{eff} .

In terms of the selection parameter recovery such a bias is undesirable. However, as we have seen above, it still comes along with a reduction of the information loss in the *statistical* model. One can therefore also ask how it affects other intrinsic properties of the orbit model. Unsurprisingly, it does not necessarily entail an improvement in these intrinsic properties either as they are also not subject to the AIC optimization. However, in practice, it often does

improve the situation. It should be noted that in general, a model's intrinsic properties (like the orbital anisotropy) are only meaningful when a priori knowledge about a galaxy's typical structure is added. In principle, any I_{mod} could be a good *statistical* model without the need of secondary properties such as inclination, orbital weights etc. (cf. Section 2). An improvement in such secondary properties can only be achieved if a property correlates with an improvement of the model LOSVDs, or in other words: if the property can be constrained by ideal (noise-free) data. As an example, the bottom panel of Fig. 24 shows the RMSDs of β analogous to the ΔI . Unlike the model LOSVDs the anisotropy recovery does not necessarily improve with the AIC model selection. For example, when α is not optimized the anisotropy recovery worsens considerably by choosing the more face-on models. However, as seen in Section 7, a substantial improvement in the orbital recovery can be made by optimizing the regularization, and in that case $\Delta\beta$ is approximately independent of the inclination, as expected. In practice, therefore, intrinsic model properties can be improved via the model selection even in a situation where parts of the model are not well constrained.

We suppose that an approach similar to our model selection can also be transferred to other modelling frameworks. For example, in a Bayesian setting it may be worthwhile to marginalize over the entire high-dimensional space of orbital weights (e.g. Magorrian 2006, 2014) and nuisance parameters (e.g. Bovy, Kawata & Hunt 2018) given some suitable prior. In such a Bayesian setting the remaining parameters which are not marginalized out, play a similar role as the selection parameters do in our model selection framework. The best set of (selection) parameters is then determined according to the resulting marginalized likelihoods of different trial models and will depend on the choice of the prior. Our results seem to make a good case that a prior choice that aims to minimize the KLD of the LOSVDs – such that the marginalized likelihoods rank models equivalently to the Akaike model selection (KLD-prior, e.g. Burnham et al. 2002) – is very powerful. It has the additional benefit of being easy to implement even for complex models.

In summary, the AIC selection will always choose the model with the least amount of effective parameters out of all models that achieved a good fit. Often this can be very advantageous, especially when determining the regularization parameter α from the data (Section 7). However, in the case of degeneracies, when the information contained in the LOSVD data is enough to constrain a selection parameter only up to some interval, then the selection *within* this degenerate interval can be biased with respect to the true values. The fact that we never encountered any related bias in our flattened toy galaxies for either i , Υ , or α suggests that all these selection parameters are *well constrained* by the kinematic observations of the kind we tested: *fully resolved LOSVDs* over a field of view typical for modern integral-field spectrographs. On top of that, secondary properties such as the anisotropy β are likewise well determined by this kind of data. This is consistent with the recent findings of Neureiter et al. (2021) in the triaxial case and shows how powerful orbit models are if all the information contained in modern data is exploited.

9.2 Inclination recovery of axisymmetric galaxies

We are not aware of any systematic attempts yet to study the inclination recovery of axisymmetric galaxies using orbit models. Early works on M32, using a χ^2 minimization, yielded promisingly good constraints ($i = 70^\circ \pm 5^\circ$, Verolme et al. 2002). However, the galaxy's intrinsic axial ratio $q = 0.68 \pm 0.03$ at this inclination is just slightly different from the $q' = 0.73$ expected for $i = 90^\circ$ and the errors were estimated using the 3σ intervals of a $\Delta\chi^2$ distribution with only three

degrees of freedom (one for each varied selection parameter). Hence, the significance of the assigned inclination uncertainty is hard to judge. A subsequent detailed modelling of the early-type galaxy NGC 2974 by Krajnović et al. (2005) yielded a similarly well constrained inclination $i = 65^\circ \pm 2.5^\circ$ according to a χ^2 -minimization. Despite the small statistical uncertainty, however, Krajnović et al. (2005) concluded that the expected kinematic differences between models at different inclinations are so small that systematics in the data and/or in the models hamper a robust inclination recovery. Modelling tests with a two-integral toy model designed to represent NGC 2974 also supported the conclusion that a determination of the inclination is likely unfeasible. These studies did not exploit all the information contained in fully resolved LOSVDs, but were based on Gauss–Hermite expansions. Noteworthy, these early studies did not exhibit an apparent edge-on bias that dominates the χ^2 surfaces. In comparison, our modelling of the toy galaxies and NGC 3368 were visibly impacted by the greater flexibility of more inclined models. Besides the galaxies/toy models presented in this paper, a dominant edge-on bias was present in the axisymmetric modelling results of NGC 4151 by Onken et al. (2007). They compared models at the most likely inclination of the galaxy's large-scale disc ($i = 23^\circ$) and at $i = 90^\circ$. The edge-on models achieved a better fit. Gebhardt et al. (2000) modelled NGC 3379 with inclinations ranging from $i = 29^\circ$ to 90° and found that the edge-on models are strongly preferred. Thomas et al. (2007a) modelled mock data derived from a suite of N -body binary merger simulations with differently inclined models yet the edge-on model consistently fitted the mocks best. Thomas et al. (2007b) modelled a sample of galaxies in the Coma cluster assuming three different inclination angles per galaxy. Most of the galaxies were found to be edge-on, but since the sample was selected towards significantly flattened galaxies a potential inclination bias was difficult to quantify. More recently Liepold et al. (2020) modelled NGC 1453 using different assumed inclinations with the triaxial Schwarzschild implementation of van den Bosch et al. (2008) in the axisymmetric limit and found that the kinematics are fitted better the closer the inclination is to an edge-on configuration.

The fact that some of the early studies did not experience a noticeable edge-on bias could be caused by discreteness effects or other systematics being more significant than the bias induced by the model flexibility. For example, in early applications of the Schwarzschild models less orbits were used than are typically used now. Alternatively, the inclination bias could also depend on the galaxy structure. For example, if a galaxy is highly flattened its intrinsic structure might involve large intrinsic azimuthal velocities (like in a rotating disc). Edge-on models might have trouble to reach such large projected net velocities due to their rounder intrinsic shapes unless the $\cos(i)$ factor in the velocity projection overcompensates the lower intrinsic rotation (at fixed mass). In most cases this would be no problem for edge-on orbit libraries as they can usually follow the rotation signal by re-balancing the weights of retrograde to prograde orbits or by increasing the mass. However, if the edge-on model is already 'at its limits', for example if all light contribution already comes from either retrograde or prograde orbits, then it may become significant. We were able to construct such extreme toy galaxies for which the edge-on bias (at fixed mass) was weaker or even absent because the edge-on models were simply not able to adequately fit the kinematic data and therefore the edge-on bias was not a dominant feature in the χ^2 surfaces. However, we regard this explanation unlikely, since these extreme toy galaxies came along with a positive (global) correlation between the projected rotation velocity v and the Gauss–Hermite parameter h_3 , in contrast to the observed anticorrelation between v and h_3 in real galaxies

(Bender et al. 1994; a positive correlation has only been observed vary rarely and even then only in small areas of a galaxy, e.g. Guérou et al. 2016).

Beyond the inclination in axisymmetric Schwarzschild implementations the viewing angles and associated shapes of triaxial systems also often prove difficult to recover correctly unless the underlying galaxy exhibits very distinct kinematic features (e.g. van den Bosch et al. 2008). By modelling simulations using triaxial Schwarzschild models Jin et al. (2019) find that the flattening of the intermediate and minor axis are on average slightly overestimated, thus, biasing the models to more spherical shapes. This bias may be responsible for the large fraction of nearly spherical systems in a sample of 149 early-type galaxies modelled triaxially by Jin et al. (2020).

Unlike Schwarzschild models isotropic/anisotropic Jeans-models often provide strong inclination constraints and kinematic maps obtained from Jeans models are noticeably different when viewed under different angles (cf. Cappellari et al. 2006; Cappellari 2008). However, these constraints may be overly optimistic as Jeans models only model the first and second velocity moments and are not as general as Schwarzschild models as they rely on assumptions about the velocity ellipsoid and anisotropy of the galaxy. The resulting Jeans models are not guaranteed to correspond to non-negative phase-space distribution functions and only a restricted subspace of all possible non-negative distribution function compatible with a given potential is sampled by Jeans models. These restrictions may be advantageous in breaking the aforementioned inclination degeneracy if the assumptions hold approximately valid for a given galaxy, but if not then Jeans modelling could generate artificially constrained inclination results.

While there do exist alternative methods to constrain the inclination that are not based on dynamical modelling they are usually only applicable to disc galaxies, meaning that the inclination and consequently the intrinsic flattening of elliptical galaxies is often undetermined. However, since viewing angles are necessary parameters in the construction of any Schwarzschild model it is often simply assumed that elliptical galaxies are viewed edge-on or one adopts the inclination derived from Jeans models for the construction of Schwarzschild models (e.g. Cappellari et al. 2006; Thater et al. 2019). A successful recovery of the inclination with dynamical modelling implies that the intrinsic 3d stellar mass distribution that corresponds to the tested viewing angles results in gravitational potentials (and thus the orbit models) that differ significantly enough to be detectable in the model selection framework. This stellar mass/luminosity distribution is obtained by deprojecting the observed surface brightness distribution for a given inclination. However, the deprojection of axisymmetric systems is ambiguous as it is only unique in the special case where the galaxy is viewed edge-on. For all other viewing angles the luminosity density is underdetermined as there exists a range of deprojections compatible with a given surface brightness distribution and inclination (cf. Rybicki 1987; Rix & White 1990). As demonstrated by Gerhard & Binney (1996) one can construct many physically reasonable boxy and discy density distributions that projected to the same surface brightness distribution for a given inclination. This means there may exist many physically reasonable deprojections that could be used for the construction of our orbit libraries at a given inclination $i < 90^\circ$. Even though we obtained the (non-parametric) deprojections used in this paper by applying an Metropolis-Algorithm (cf. Magorrian 1999) which is able to explore the full range of physically allowed deprojections, we only picked a single ‘privileged’ deprojection per assumed inclination, as is common in modelling applications to real galaxies. Given our results are suggesting that it is possible to distinguish between

different inclinations and the differently flattened stellar distributions associated with them, it may also be possible to further dynamically discriminate between the many boxy/discy deprojections compatible with a given surface brightness and inclination. This possibility of breaking the (photometric) deprojection degeneracy using kinematics was explored before by Magorrian (1999) who has shown for *two-integral* models that discs which are undetectable in the photometry can leave easily detectable features in the corresponding kinematics. We plan to investigate the deprojection degeneracy in the future using our model selection framework.

9.3 Mass reconstruction

Similar to the inclination other selection parameters can be biased, in particular mass parameters. For example we found the mass-to-light ratios of the toy galaxies A and D to be biased high by about 3–4 per cent when evaluated with a χ^2 minimization because their intrinsic model flexibility is positively correlated with Υ . In addition to this explicit dependence on the m_{eff} we also expect that mass parameters are implicitly affected if the galaxy is modelled assuming the wrong viewing angles. Thomas et al. (2007a) discuss via the tensor virial theorem how this depends on the intrinsic shape of a galaxy. For example, the masses of oblate objects seen face-on are typically underestimated, while the masses of prolate objects seen end-on will be overestimated. In the triaxial modelling of the Milky Way’s nuclear star, Feldmeier-Krause et al. (2017) find that the mass-to-light ratio is positively correlated with the triaxial shape parameter p (or the equivalent viewing angle) while the black hole mass only varies slightly. However, the χ^2 surfaces for both mass parameters broaden as said shape parameter is increased. In the axisymmetric case the dynamical modelling of NGC 4151 by Onken et al. (2007) indicates a strong correlation of both, black hole mass and mass-to-light ratio, with the assumed inclination. Our modelling of NGC 3368 only very tentatively exhibits such a correlation between the two mass parameters M_\bullet and Υ with the assumed inclination (Fig. 19), however, we do notice an increased scatter when modelling the galaxy edge-on. We generally expect that mass parameters which are less constrained by the available kinematic data (often these are the parameters of a dark matter halo or the mass of a central SMBH) are more sensitive to varying model flexibilities and, thus, are more likely prone to a bias. In multicomponent mass models an additional complication arises from the cross-talk between the various components (e.g. Erwin et al. 2018). Whether a specific component could be over- or underestimated is therefore not entirely clear. A more focused and systematic investigation with a large galaxy sample or with multicomponent toy models could give better insight. We plan to address these questions and also the triaxial case in future papers.

10 SUMMARY AND CONCLUSIONS

When modelling the kinematic data of a galaxy to determine the mass of its SMBH, the stellar initial mass function or the structure of its dark matter halo, a huge number of trial dynamical models with different assumed mass distributions have to be fitted to the data. Then, based on the observations and the quality of the fit, one needs to decide which of these models represents the true structure of the galaxy best. We have motivated that the commonly used approach to judge the models solely by their goodness-of-fit is often not well defined. Moreover, we have shown that it can lead to substantial biases in estimated galaxy properties. The reason is that the process of identifying the best fit involves the comparison of *different*

models with intrinsically different model flexibilities. In the case of axisymmetric modelling this point is most apparent when trying to recover a galaxy’s inclination, where it causes a dominant bias towards edge-on models. If not corrected for, this effect inhibits the possibility of constraining the inclination via dynamical modelling. However, the issue of varying model flexibility is not contained to the inclination alone, we demonstrated that it also introduces an overestimation of galaxy masses.

Quantifying the flexibility of Schwarzschild models is non-trivial due to various complications like nonlinearity or priors that restrict the parameter space accessible to the orbital weights w_i . We introduced the concept of the number of effective free parameters and presented two calculation methods which rely on bootstrap iterations. Although computationally expensive, this bootstrap approach is robust, flexible and can be applied to more general modelling techniques. Once the model flexibility (i.e. the number of effective parameters m_{eff}) is known, model selection techniques can be applied to choose the best model out of a given set of fits.

We tested three different model selection frameworks, the ‘intuitive’ one, an information-based one using the Akaike Information Criterion (AIC) and a more generalized approach with an adjustable parameter w_m . The only difference between the approaches is how the model flexibility is weighted in the model selection process. The ‘intuitive’ approach considers models equivalent that have the same reduced χ^2 ($\chi^2 + m_{\text{eff}} \rightarrow \min$), the AIC weighs the effective parameters twice as large as in the intuitive approach ($\chi^2 + 2m_{\text{eff}} \rightarrow \min$) and yields the least flexible model out of all models with the same reduced χ^2 . In the generalized approach ($\chi^2 + w_m m_{\text{eff}} \rightarrow \min$) the weight of the model flexibility is parametrized by w_m . It includes the other two approaches as special cases.

We applied these model selection schemes to realistic mock data sets of a number of axisymmetric toy galaxies with the goal to recover their inclination and mass-to-light ratio Υ . We confirmed that an evaluation based solely on χ^2 always favours the edge-on orbit models and is biased towards higher mass-to-light ratios. A model assessment based on χ^2 alone will limit the potential constraining power of dynamical models, meaning that better kinematic data will not lead to a corresponding improved accuracy in the estimated galaxy properties. Model evaluation within a model selection framework can correct these issues, enabling the recovery of the correct galaxy inclination and mass with very small uncertainties.

In a second step we extended the model selection approach to also encompass the strength of the entropy regularization that is applied in the models. The amount of regularization is a crucial choice as it can negatively affect the recovery of the intrinsic dynamics and the stability of the calibration weight w_m . This led to a further significant improvement of the results. Based on the simulated toy galaxies we found that

- (i) the model selection not only returns always the correct masses and inclinations but also returns the model that matches the toy galaxy most closely in terms of its orbital dynamics;
- (ii) while the best results of the model selection with fixed smoothing required a $w_m \approx 1.5$ depending on the galaxy under study, the best results with optimized smoothing were *robustly* obtained always for the *same* $w_m = 2$ (AIC);
- (iii) the constraining power of the data improved, i.e. the confidence regions of the derived galaxy masses, inclinations and orbital anisotropies tightened significantly.

This suggests that, in order to achieve optimal results, one should construct models by varying the degree of regularization among other

parameters and evaluate models with different selection parameters using the information-based Akaike criterion.

With modern integral-field spectrographs on 10 m-class telescopes it is possible to measure the LOSVDs of galaxies non-parametrically at hundreds of positions spread over the galaxy on the sky (e.g. Mehran et al. 2019). Our orbit models are designed to deal with the full amount of information that is contained in these non-parametric LOSVD fields. With the simulated toy galaxies that we used to mimic such observations in a realistic fashion we could show that orbit models allow to reconstruct galaxy inclinations, masses and anisotropies with an uncertainty no larger than 1-2 per cent in mass and $\Delta\beta \sim 0.04$ in the anisotropy. As long as the regularization is optimized the viewing angle and consequently the intrinsic flattening q of the toy galaxies are well constrained by an AIC framework with an uncertainty smaller than the sampling size $\Delta i = 10^\circ$ of the trial models used here. This demonstrates the power of orbit superposition models and the prospects of the model selection ansatz.

The edge-on bias in the χ^2 -minimization is also occurring for the dynamical modelling of real galaxies, as demonstrated by the modelling of the spiral galaxy NGC 3368. Even though the ellipticity of the galaxy’s large-scale disc suggests a moderate inclination of $i = 53^\circ$, edge-on models fit kinematic data obtained with SINFONI systematically better than dynamical models constructed with $i = 53^\circ$. When the varying flexibilities are included, however, the recovered inclination is in agreement with the observed ellipticity. Specifically, when applying the full model selection with optimized regularization, we find a black hole mass $M_\bullet = (5.5 \pm 3.3) \times 10^6 M_\odot$, a bulge mass-to-light ratio $\Upsilon_{\text{bulge}} = (0.55 \pm 0.11) M_\odot/L_\odot$ and an inclination angle $i = (57.8 \pm 5.1)^\circ$ in agreement with the independent inclination estimation. The fact that we could recover the galaxy’s inclination from just the SINFONI kinematics (which cover only the inner ~ 1 arcsec of the galaxy, roughly 5-10 times the sphere of influence of the central SMBH) again underlines the prospect of the model selection technique. The above Monte Carlo simulations suggest that the accuracy of the involved selection parameters can be significantly improved by modelling more extended, high-resolution kinematic data. In addition, we suspect that a modelling that also takes into account more accurate, correlated error patterns will further improve the dynamical modelling. Thus, the obvious next step will be to incorporate the full error correlation matrix of the observed LOSVDs within the modelling procedure.

The model selection approach is versatile and may well be used to optimize other intrinsic library parameters which impact the model flexibility like, e.g. the number of orbits. We plan to investigate this in a companion paper. Since the objective of the AIC selection is to minimize the KLD of the *statistical* model (i.e. in our case the LOSVDs) it does not rely on the internal structure of the underlying model. Therefore the approach may be adopted for other dynamical modelling techniques as well. It may even be possible that the extension of the model selection ansatz to determine the optimal smoothing based on data can be applied in other non-parametric methods as well, e.g. in the recovery of non-parametric LOSVDs (Thomas et al., in preparation), non-parametric deprojections (e.g. de Nicola et al. 2020) and non-parametric source reconstructions in strong gravitational lensing, etc.

ACKNOWLEDGEMENTS

We are grateful to Dr. Roberto Saglia for his active support and for many valuable discussions throughout the project. We also thank the anonymous referee for helpful comments. This research was supported by the Excellence Cluster ORIGINS which is funded

by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094-390783311. Simulations have been carried out on the computing facilities of the Computational Center for Particle and Astrophysics (C2PAP) and at the COBRA HPC system at the Max Planck Computing and Data Facility (MPCDF), Germany.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Andrae R., Schulze-Hartung T., Melchior P., 2010, preprint ([arXiv:1012.3754](https://arxiv.org/abs/1012.3754))
- Baes M., Dejonghe H., Buyle P., 2005, *A&A*, 432, 411
- Bender R., Saglia R. P., Gerhard O. E., 1994, *MNRAS*, 269, 785
- Binney J., Tremaine S., 2008, *Galactic Dynamics*, 2nd edn. Princeton Univ. Press, Princeton
- Bovy J., Kawata D., Hunt J. A. S., 2018, *MNRAS*, 473, 2288
- Burnham K. P., Anderson D. R., Burnham K. P., 2002, *Model Selection and Multi-model Inference: A Practical Information-theoretic Approach*. Springer, New York
- Cappellari M., 2008, *MNRAS*, 390, 71
- Cappellari M., Copin Y., 2003, *MNRAS*, 342, 345
- Cappellari M., Verolme E. K., van der Marel R. P., Verdoes Kleijn G. A., Illingworth G. D., Franx M., Carollo C. M., de Zeeuw P. T., 2002, *ApJ*, 578, 787
- Cappellari M. et al., 2006, *MNRAS*, 366, 1126
- Cretton N., Emsellem E., 2004, *MNRAS*, 347, L31
- Cretton N., de Zeeuw P. T., van der Marel R. P., Rix H.-W., 1999, *ApJS*, 124, 383
- de Nicola S., Saglia R. P., Thomas J., Dehnen W., Bender R., 2020, *MNRAS*, 496, 3076
- Emsellem E. et al., 2011, *MNRAS*, 414, 888
- Erwin P., 2004, *A&A*, 415, 941
- Erwin P., Thomas J., Saglia R. P., Fabricius M., Rusli S. P., Seitz S., Bender R., 2018, *MNRAS*, 473, 2251
- Feldmeier-Krause A., Zhu L., Neumayer N., van de Ven G., de Zeeuw P. T., Schödel R., 2017, *MNRAS*, 466, 4040
- Gebhardt K. et al., 2000, *AJ*, 119, 1157
- Gebhardt K. et al., 2003, *ApJ*, 583, 92
- Gerhard O. E., 1993, *MNRAS*, 265, 213
- Gerhard O. E., Binney J. J., 1996, *MNRAS*, 279, 993
- Guérou A., Emsellem E., Krajnović D., McDermid R. M., Contini T., Weilbacher P. M., 2016, *A&A*, 591, A143
- Häfner R., Evans N. W., Dehnen W., Binney J., 2000, *MNRAS*, 314, 433
- Hastie T., Tibshirani R., Friedman J., 2013, *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Springer Science & Business Media, Berlin
- Hernquist L., 1990, *ApJ*, 356, 359
- Ho L. C., Filippenko A. V., Sargent W. L. W., 1997, *ApJS*, 112, 315
- Houghton R. C. W., Magorrian J., Sarzi M., Thatte N., Davies R. L., Krajnović D., 2006, *MNRAS*, 367, 2
- Jin Y., Zhu L., Long R. J., Mao S., Xu D., Li H., van de Ven G., 2019, *MNRAS*, 486, 4753
- Jin Y., Zhu L., Long R. J., Mao S., Wang L., van de Ven G., 2020, *MNRAS*, 491, 1690
- Knapen J. H., de Jong R. S., Stedman S., Bramich D. M., 2003, *MNRAS*, 344, 527
- Kowalczyk K., Łokas E. L., Valluri M., 2017, *MNRAS*, 470, 3959
- Krajnović D., Cappellari M., Emsellem E., McDermid R. M., de Zeeuw P. T., 2005, *MNRAS*, 357, 1113
- Krajnović D., McDermid R. M., Cappellari M., Davies R. L., 2009, *MNRAS*, 399, 1839
- Leung G. Y. C. et al., 2018, *MNRAS*, 477, 254
- Levison H. F., Richstone D. O., 1985, *ApJ*, 295, 349
- Liebold C. M., Quenneville M. E., Ma C.-P., Walsh J. L., McConnell N. J., Greene J. E., Blakeslee J. P., 2020, *ApJ*, 891, 4
- Magorrian J., 1999, *MNRAS*, 302, 530
- Magorrian J., 2006, *MNRAS*, 373, 425
- Magorrian J., 2014, *MNRAS*, 437, 2230
- Mehrgan K., Thomas J., Saglia R., Mazzalay X., Erwin P., Bender R., Kluge M., Fabricius M., 2019, *ApJ*, 887, 195
- Morganti L., Gerhard O., 2012, *MNRAS*, 422, 1571
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Neureiter B. et al., 2021, *MNRAS*, 500, 1437
- Nowak N., Thomas J., Erwin P., Saglia R. P., Bender R., Davies R. I., 2010, *MNRAS*, 403, 646
- Onken C. A. et al., 2007, *ApJ*, 670, 105
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical recipes in FORTRAN: The Art of Scientific Computing*. Cambridge Univ. Press, New York, NY
- Richstone D. O., Tremaine S., 1988, *ApJ*, 327, 82
- Rix H.-W., White S. D. M., 1990, *ApJ*, 362, 52
- Rix H.-W., de Zeeuw P. T., Cretton N., van der Marel R. P., Carollo C. M., 1997, *ApJ*, 488, 702
- Rusli S. P. et al., 2013, *AJ*, 146, 45
- Rybicki G. B., 1987, in de Zeeuw P. T., ed., *Proc. IAU Symp. Vol. 127, Structure and Dynamics of Elliptical Galaxies*. Kluwer, Dordrecht, p. 397
- Saglia R. P., Kronawitter A., Gerhard O., Bender R., 2000, *AJ*, 119, 153
- Saglia R. P. et al., 2016, *ApJ*, 818, 47
- Schwarzschild M., 1979, *ApJ*, 232, 236
- Siopis C., Kandrup H. E., 2000, *MNRAS*, 319, 43
- Thater S., Krajnović D., Cappellari M., Davis T. A., de Zeeuw P. T., McDermid R. M., Sarzi M., 2019, *A&A*, 625, A62
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Richstone D., 2004, *MNRAS*, 353, 391
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2005, *MNRAS*, 360, 1355
- Thomas J., Jesseit R., Naab T., Saglia R. P., Burkert A., Bender R., 2007a, *MNRAS*, 381, 1672
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2007b, *MNRAS*, 382, 657
- Thomas J. et al., 2009a, *MNRAS*, 393, 641
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2009b, *ApJ*, 691, 770
- Thomas J. et al., 2011, *MNRAS*, 415, 545
- Thomas J., Saglia R. P., Bender R., Erwin P., Fabricius M., 2014, *ApJ*, 782, 39
- Tony J. L., Dressler A., Blakeslee J. P., Ajhar E. A., Fletcher A. B., Luppino G. A., Metzger M. R., Moore C. B., 2001, *ApJ*, 546, 681
- Valluri M., Merritt D., Emsellem E., 2004, *ApJ*, 602, 66
- van de Ven G., de Zeeuw P. T., van den Bosch R. C. E., 2008, *MNRAS*, 385, 614
- van den Bosch R. C. E., van de Ven G., Verolme E. K., Cappellari M., de Zeeuw P. T., 2008, *MNRAS*, 385, 647
- van der Marel R. P., Franx M., 1993, *ApJ*, 407, 525
- van der Marel R. P., Cretton N., de Zeeuw P. T., Rix H.-W., 1998, *ApJ*, 493, 613
- Vasiliev E., Athanassoula E., 2015, *MNRAS*, 450, 2842
- Vasiliev E., Valluri M., 2020, *ApJ*, 889, 39
- Verolme E. K. et al., 2002, *MNRAS*, 335, 517
- Ye J., 1998, *J. Am. Stat. Assoc.*, 93, 120
- York D. G. et al., 2000, *AJ*, 120, 1579

APPENDIX A: BASICS - CONSTRUCTING SCHWARZSCHILD MODELS

To derive the orbital structure of a galaxy or the mass of its central supermassive black hole, stars and dark matter halo via dynamical modelling, one starts from a set of photometric and kinematic

observations. The photometric observations typically consist of measurements of a galaxy’s surface brightness either from an isophote analysis or directly from an image. Even though the photometric information is intrinsically 2d in nature, we can always rearrange the finite number of measurements into a 1d vector $\mathbf{s}b_{\text{obs},i}$ with $i = 1, \dots, N_{\text{phot}}$ data points and their corresponding errors $\Delta \mathbf{s}b_{\text{obs},i}$. The kinematic observations consist of a large number of measurements related to the LOSVDs $\mathbf{l}_{\text{obs},ij} \pm \Delta \mathbf{l}_{\text{obs},ij}$ at $i = 1, \dots, N_{\text{kin}}$ positions spread over the galaxy on the sky (e.g. in N_{kin} Voronoi Bins). If the LOSVDs are measured non-parametrically, which is the state of the art today (e.g. Mehrgan et al. 2019), then in each bin on the sky we have $j = 1, \dots, N_{\text{vel}}$ data points that sample the respective LOSVD over N_{vel} different line-of-sight velocities. For modern integral-field spectrographs, the $\mathbf{l}_{\text{obs},ij}$ form a data cube and are 3d in nature. Still, we can again rearrange the finite (though large) number of data points into a 1d vector $\mathbf{l}_{\text{obs},k}$ or short \mathbf{l}_{obs} . Several modelling steps are required to derive the mass distribution or internal structure of a galaxy from an orbit superposition model based on these observations \mathbf{l}_{obs} and $\mathbf{s}b_{\text{obs}}$.

First, a trial mass distribution has to be assumed. The corresponding gravitational potential is calculated via the Poisson equation. Then a set of orbits with representative initial conditions, called the orbit library, is integrated in this potential and the properties of all orbits are stored. It is necessary to sample the available phase-space sufficiently dense with these orbits, thus their initial conditions must be chosen with care. For a comprehensive description of our orbit sampling, see Thomas et al. (2004). The orbits of the library are then superimposed by giving each of them an adjustable, non-negative orbital weight w_i , akin to an occupation number that represents the number of stars tracking the orbit. The non-negativity constraint is imposed on these orbital weights to guarantee the resulting phase-space distribution function of the orbit model is positive everywhere. The adjustment of the orbital weights is done such that the properties of the superposition emulate the photometric and kinematic observations of the galaxy as good as possible in the assumed trial potential. Here it turns out that the LOSVDs $\mathbf{l}_{\text{mod},j}$ of the orbit superposition model are linear combinations of the individual contributions of each orbit. In other words, if the contribution of orbit i to the kinematic measurement j is $\mathbf{l}_{\text{orb},i}^j$, then the predicted kinematics $\mathbf{l}_{\text{mod},j}$ of the whole orbit superposition model reads

$$\mathbf{l}_{\text{mod},j} = \sum_i^{N_{\text{orbit}}} w_i \mathbf{l}_{\text{orb},i}^j \quad (\text{A1})$$

where the sum goes over the number of library orbits N_{orbit} . In compact matrix notation we can write $\mathbf{l}_{\text{mod}} = \mathbf{L}_{\text{orb}} \cdot \mathbf{w}$, where \mathbf{L}_{orb} is the matrix with elements $\mathbf{l}_{\text{orb},i}^j$ and \mathbf{w} is the vector with the orbital weights.

The 3d mass distribution ρ required in the first step is usually unknown. A comprehensive trial mass distribution should include the most important galaxy components. Commonly, the density is composed as

$$\rho = \Upsilon \cdot \nu + M_{\bullet} \cdot \frac{\delta(r)}{4\pi r^2} + \rho_{\text{DM}} \quad (\text{A2})$$

where the first term is the contribution of the stellar component determined by the mass-to-light ratio Υ and the 3d luminosity density ν . The second term is a central supermassive black hole with mass M_{\bullet} and the third term encompasses the contribution of the dark matter halo, which can in itself be further parametrized, for example by a Navarro–Frank–White profile (Navarro, Frenk & White 1996). The 3d luminosity density ν is typically not a free parameter of this

mass model. Instead it is calculated by deprojecting the 2D surface brightness distribution of the investigated galaxy, implying that ν and consequently ρ depend on the galaxy’s assumed inclination. Similarly as for the model LOSVDs it turns out that its intrinsic 3d luminosity density \mathbf{d}_{mod} is simply the linear combination of the individual orbital contributions. In compact matrix notation we can write $\mathbf{d}_{\text{mod}} = \mathbf{D}_{\text{orb}} \cdot \mathbf{w}$, analogously to the model’s LOSVDs.

In detail, the few implementations of the Schwarzschild method that have been developed differ considerably. Some, like ours, exploit the information contained in the entire (though binned) LOSVDs of the galaxy and orbit model, while others rely on Gauss–Hermite expansions up to some finite Hermite order n (often $n < 8$). Likewise, some implementations fit the observed surface brightness and/or the deprojected density, while others – like ours – enforce full consistency of the models through equality constraints for the deprojected luminosity density \mathbf{d}_{data} . Specifically, for our implementation, we use a χ^2 minimization to derive the best-fit orbital weights from the \mathbf{l}_{obs} where the orbital weights must fulfil the equality constraint $\mathbf{d}_{\text{data}} = \mathbf{d}_{\text{mod}} = \mathbf{D}_{\text{orb}} \cdot \mathbf{w}$ to ensure self-consistency.

When modelling a galaxy, it is often one of the main goals to determine its unknown mass distribution while the orbital weights are not of primary interest. Therefore one typically postulates a number of trial mass distributions and creates an orbit superposition for each of them. Then, the final task is to pick up the ‘best’ model out of this set of trial mass distributions. Usually, this is also done via a χ^2 comparison, which means that the model with the smallest χ^2 is considered to be the best representation of the galaxy.

APPENDIX B: EXAMPLES FOR KINEMATIC MAPS

Shown are the Gauss–Hermite coefficients up to h_4 of the LOSVDs of the spherical N -body (Figs B1–B2) and of toy galaxy D (Fig. B3) which was created using Schwarzschild models with an angular momentum bias $\lambda = 0.5$.

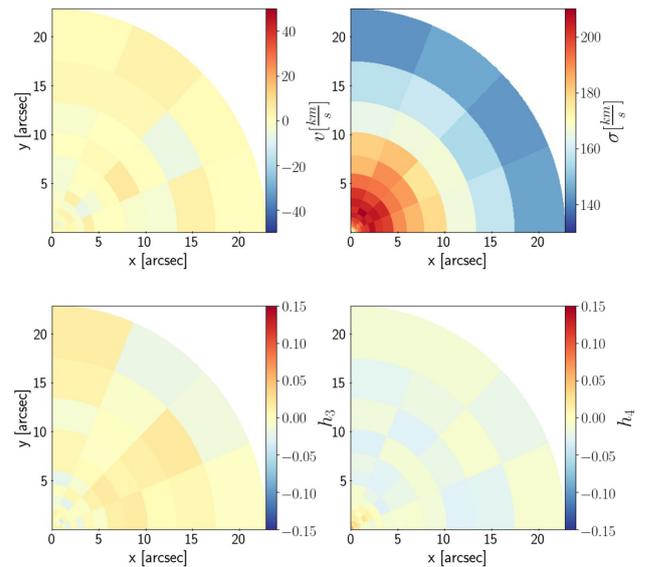


Figure B1. The mock Gauss–Hermite maps of a quadrant of the spherical Hernquist N -body. Responsible for the deviations from spherical symmetry is Gaussian distributed Monte Carlo noise that has been added to the N -body data.

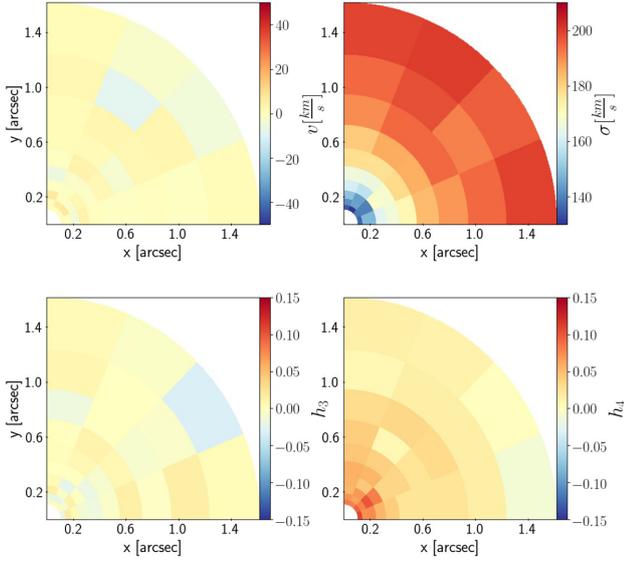


Figure B2. As Fig. B1 but only the central 2 arcsec are shown. Towards the centre the velocity dispersion drops while h_4 increases (as expected for an isotropic Hernquist galaxy, cf. Baes et al. 2005).

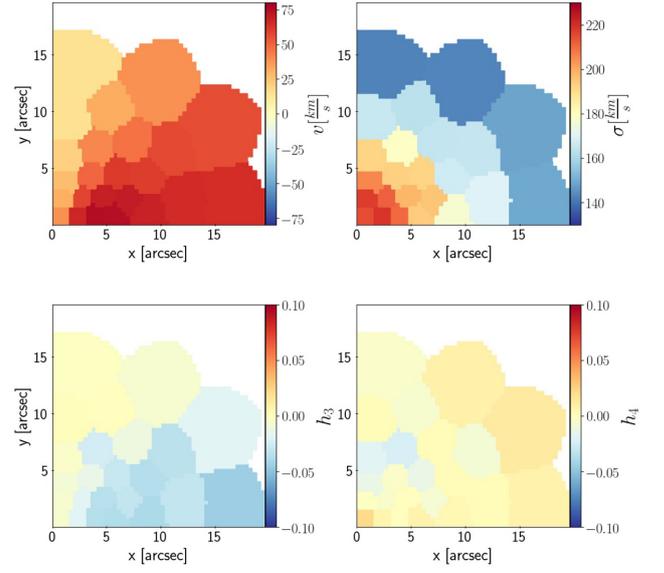


Figure B3. The kinematic map of the early-type toy galaxy D before the addition of Monte Carlo noise. The spatial grid shown is typical for wide-field kinematic data that extend beyond the galaxy’s effective radius. In addition to the spatial bins shown here we simultaneously modelled bins with higher resolution in the centre of the galaxy (cf. Fig. 8).

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

Chapter 3

The VIRUS-dE Survey I: Stars in dwarf elliptical galaxies - 3D dynamics and radially resolved stellar initial mass functions

Bibliographic information

Lipka Mathias, Jens Thomas, Roberto Saglia, Ralf Bender, Maximilian Fabricius, Gary J. Hill, Matthias Kluge, Martin Landriau, Ximena Mazzalay, Eva Noyola, Taniya Parikh, Jan Snigula - The VIRUS-dE Survey I: Stars in dwarf elliptical galaxies - 3D dynamics and radially resolved stellar initial mass functions, *The Astrophysical Journal*, Volume TBA, Issue TBA, pp.???-???
DOI: TBA

The Author's contribution

The VIRUS-W observations were obtained by Maximilian Fabricius, Gary Hill, Matthias Kluge, Martin Landriau, Ximena Mazzalay, Eva Noyola, and Jan Snigula. I reduced the raw VIRUS-W images myself, following the helpful instructions by Max Fabricius. For all galaxies I binned and extracted the stellar kinematics from the resulting data cubes using WINGFIT and performed the orbit superposition modelling with LIBMOD. Both codes were adapted to the project and provided by Jens Thomas. The single stellar population modelling of the data cubes was performed by Taniya Parikh. For all, but VCC1261 and VCC 2048, I obtained the 3D luminosity distributions using HST archival data. I reduced them, fitted isophotes and deprojected the resulting photometry for all galaxies. As first author, I wrote all sections and produced all figures in the manuscript. Jens Thomas provided guidance and extensive support to this work in the form of numerous discussions, suggestions, and text revisions. Roberto Saglia and Ralf Bender provided

further corrections and suggestions.

Copyright

©2024 MNRAS
reprinted on pages 67- 122

The VIRUS-dE Survey I: Stars in dwarf elliptical galaxies - 3D dynamics and radially resolved stellar initial mass functions

MATHIAS LIPKA ^{1,2} JENS THOMAS ^{1,2} ROBERTO SAGLIA ^{1,2} RALF BENDER ^{1,2} MAXIMILIAN FABRICIUS ^{1,2}
GARY J. HILL ^{3,4} MATTHIAS KLUGE ^{1,2} MARTIN LANDRIAU ⁵ XIMENA MAZZALAY,¹ EVA NOYOLA,⁴
TANIYA PARIKH ¹ AND JAN SNIGULA ^{1,2}

¹*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching*

²*Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany*

³*McDonald Observatory, The University of Texas at Austin, 2515 Speedway Boulevard, Stop C1402, Austin, TX 78712, USA*

⁴*Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Stop C1400, Austin, TX 78712, USA*

⁵*Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley 94720, CA, USA*

Submitted to ApJ

ABSTRACT

We analyse the stellar structure of a sample of dwarf ellipticals (dE) inhabiting various environments within the Virgo cluster. Integral-field observations with a high spectral resolution allow us to robustly determine their low velocity dispersions ($\sim 25 \text{ km s}^{-1}$) and higher-order kinematic moments out to the half-light radius. We find the dEs exhibit a diversity in ages with the younger dEs being less enhanced than the older, suggesting a complex star formation history for those dEs that recently entered Virgo while others have been quenched shortly after reionization. Orbit-superposition modeling allowed us to recover viewing angles, stellar mass-to-light ratios (with gradients), as well as the intrinsic orbit structure. We find that the angular momentum of the dEs is strongly suppressed compared to ordinary early-type galaxies and correlates with the environment. Flattened dEs are so because of a suppressed kinetic energy perpendicular to their equatorial plane. Combining population and dynamical modeling results, we find an age-dependent stellar initial mass function (IMF) or, alternatively, evidence for a more extended star formation history for those galaxies that have had higher initial mass and/or inhabited lower density environments. dEs appear to have a spatially homogeneous stellar structure but the state they were ‘frozen’ in as they stopped forming stars varies dramatically according to their initial conditions.

Keywords: Galaxy structure(622) — Galaxy formation(595) — Dwarf elliptical galaxies(415) — Virgo Cluster(1772) — Stellar kinematics(1608)

1. INTRODUCTION

Dwarf galaxies with $\log_{10}(L_B/L_\odot) \in [8 - 10]$ dominate the galaxy census in the local Universe (Ohlson et al. 2024). Within dense galaxy clusters, like Virgo or Fornax, the *quiescent* dwarf elliptical galaxies (dE) in this luminosity range are the most abundant type of dwarfs (Sandage et al. 1985; Paudel et al. 2023). Still their origin, evolution, cosmological role and relation to

other galaxy classes still poses many questions. Like their namesakes, the more massive elliptical galaxies, dEs appear to be well approximated by featureless ellipsoids without any substantial substructures which have ceased star formation lacking sufficient gas reservoirs to form new stars. Furthermore, similar to these giant early-type galaxies (ETGs) (Dressler 1980), the dwarf early-types seem to follow a strong morphology–environment dichotomy: dEs are found predominately in dense cluster and group environments with fewer dEs interspersed in the field. In contrast, late-type dwarfs of comparable mass avoid these denser environments and

inhabit almost exclusively lower density field environments (Binggeli et al. 1987; Geha et al. 2012).

Substantial evidence suggests that dEs ($M_B \gtrsim -18$) are physically distinct from the brighter giant ETGs and should be thought of as a separate galaxy class with its own unique formation channel. For example, dEs differ in their location on the fundamental plane (Bender et al. 1992), have almost exponential light profiles (Kormendy 1985), and when investigated in detail often exhibit substructure such as faint spiral arms, blue centers, or signs of tidal harassment (Jerjen et al. 2000; Lisker et al. 2006b, 2007; Paudel & Ree 2014). In this context dEs appear to be much more closely related to the smallest galaxies in the ETG sequence: the dwarf spheroidal galaxies (dSphs) with $M_B \gtrsim -13$ which are found as satellites of the Milky Way and M31 in the Local Group.

These discrepancies to the giant galaxies in the ETG sequence suggests a distinct formation scenario for dwarf galaxies: dEs similar to their even less massive associates dSphs, are believed to be the remnants of transformed late-type galaxies (LTG), which have lost most gas through quenching (Kormendy et al. 2009). In this context some of the dEs may be better classified as Sphs which form the extension of the S0-branch in parallel to their irregular dwarf progenitors (Kormendy & Bender 2012).

This transformation from late-type to early-type may have happened through a combination of processes induced by interactions with the local environment such as ram-pressure stripping (Gunn & Gott 1972; Lin & Faber 1983), starvation (Larson et al. 1980), and galaxy harassment (Moore et al. 1998), while some of the smallest dwarf galaxies may even directly originate from tidal interactions of more massive galaxies (Barnes & Hernquist 1992; Yang et al. 2014). Which and by how much these processes are important is still a point of contention as it is difficult to disentangle their effects. Furthermore, the driving quenching mechanism may vary with the total mass, age and environment.

To improve our understanding of where, when, and how dEs have formed their *intrinsic* properties need to be studied in more detail. For example, little is known about the orbital composition of dwarf elliptical galaxies. Do they follow a similar correlation between flattening and anisotropy as intermediate-mass or very massive early-type galaxies do (Cappellari et al. 2007; Thomas et al. 2009b; Santucci et al. 2022)? Likewise, we do not know much about the dark-matter (DM) content of these galaxies. Do they follow the scaling relations of disk galaxies (Kormendy & Freeman 2016) or do they have denser DM halos, similar to ETGs (Gerhard et al.

2001; Thomas et al. 2009a)? Finally, new insight might come from studying the shape of the initial stellar mass function (IMF). In particular, comparison of the IMF of dEs to that of LTGs could shed light on the question whether the star-formation conditions in dEs and LTGs were different from the beginning or whether their evolutionary paths diverged only later after most of the stars had already formed.

This study is part of a pair of papers with the overarching goal to recover the intrinsic mass and kinematic structure of the dEs, investigate their formation, evolution, and relation to other galaxy classes and what it implies for cosmological structure formation in a broader context. We have targeted a sample of dEs in the intermediate magnitude range $M_B \sim -17$, i.e. fainter than giant ETGs but still brighter than dSphs. All of them are part of the Virgo cluster, but inhabit locally different environments: ranging from the very center, over subclumps, to the outskirts of the cluster. Primarily, we analysed spatially resolved spectra obtained for these dEs. Similar spectroscopic studies that are focused on cluster dEs, studying their stellar population properties and projected velocities and dispersions, have been conducted in the past (e.g. Geha et al. 2002; van Zee et al. 2004; Paudel et al. 2010; Ryś et al. 2013; Toloba et al. 2014; Scott et al. 2020; Bidaran et al. 2020). However, due to the low velocity dispersions and low surface brightnesses of dEs previous spectra often had too low resolution, covered the kinematics only along slits, or had insufficient signal-to-noise to infer higher-order kinematic moments beyond the rotation velocity and velocity dispersion. Here, instead, we analyse spectra obtained with the VIRUS-W spectrograph which achieves sufficient spectral resolution ($R = 7900 - 9000$) and is able to exploit the full 2D kinematic information available on the sky. The high signal-to-noise of our data allows us to study the higher-order kinematic moments beyond rotation and velocity dispersion in an unbiased manner. This novel information is a requirement to go beyond an analysis of the on-sky structure of the dEs and infer their 3D structure with dynamical models. We use sophisticated, orbit-based dynamical models to recover the intrinsic structure of the galaxies (e.g. dark matter halos, 3D orbit structure, black holes).

The current study presents the observational data we obtained and outlines the information extraction techniques we applied. In the second half of the paper we present the first results which are focused on the *stellar* component of the dEs. Among other things we examine their stellar mass, stellar mass-to-light ratio gradients, projected and intrinsic kinematic structure, ages and metallicities as well as the form of the IMF. In a com-

panion paper (in the following VW–II) we will present the corresponding *dark* components of the dEs, i.e. the dark matter halos and black holes, and discuss the implications of our results in the broader context of Λ CDM cosmology and galaxy evolution.

The current paper is organized as follows: In Section 2 we present the photometric and spectroscopic data sets and describe how we processed them to obtain the input for the dynamical models: the 3D luminosity distribution ν and the spatially resolved line of sight velocity distributions (LOSVDs). We explain our SSP modeling approach and we outline the dynamical modeling technique we employed, the Schwarzschild (1979) method, and motivate our choice of its implementation and sampling strategy. Then in Section 3 we present the on-sky kinematic structure and SSP properties and investigate their relation to the cluster environment. The 3D stellar mass, mass-to-light ratio gradients, and intrinsic anisotropy structure that was inferred using the dynamical models are presented in Section 4. Finally, in Sec. 5 we use both the dynamical and SSP constraints in combination to discuss their physical implications regarding their IMF and star-formation history. The paper concludes with a summary in Section 6. In App. A we show an example of a typical LOSVD recovery from a VIRUS-W spectrum. We compile and compare existing kinematic and SSP results from the literature with our results in App. B. In App. C and App. D we discuss the robustness of the mass estimates from the SSP models and show a supplementary IMF-metallicity relation. In App. E we discuss each dE individually based on the photometric and kinematic data we obtained. Two alternative Figures to those in Section 3 are located in App. F.

2. GALAXY SAMPLE AND DATA EXTRACTION

We investigate a total of 9 dwarf elliptical (dE) galaxies located in the Virgo Cluster which occupy the small apparent magnitude range of $m_B \in [15.0, 13.5]$ mag or in absolute magnitudes $M_B \in [-16.3, -17.8]$ mag. Fig. 1 is a thumbnail gallery of the dE galaxies obtained from the g, r, i images captured by the Sloan Digital Sky Survey (Ahn et al. 2012). We follow the general dE classification of Lisker et al. (2007), i.e. galaxies that are sometimes classified as dS0 or have faint spiral arms are also included in the dE class. The galaxies were chosen such that (i) the sample includes as many sub-classes of dEs as possible and (ii) different environments within Virgo are probed. Specifically, the sample includes the more common nucleated dE(N) and non-nucleated dE(nN) and the less common blue-centered dE(bc) and disk dE(di) (Lisker et al. 2007). This

choice was done purposefully with the goal to identify possibly different formation and evolution mechanisms that may correlate with environment or morphological substructure. Notably missing are visibly merging or tidally disrupted dEs which are much rarer (Paudel et al. 2023). The sample dEs explore a range of g -band surface brightnesses from $\mu_e \sim 22.2$ mag/arcsec² down to $\mu_e \sim 23.3$ mag/arcsec² measured at the stellar effective radius r_e . The galaxy sizes range from $r_e = 0.8$ kpc to $r_e = 1.6$ kpc (e.g. Ferrarese et al. 2006) which is fairly representative of dEs in our magnitude range. Our sample covers mostly the brighter end of the dE distribution (e.g. Kormendy & Bender 2012; Paudel et al. 2023), missing fainter and more diffuse dEs, and, notably, does not contain the even fainter ultra-diffuse galaxies (UDGs) which sometimes are regarded as their own subclass.

Table 1 lists some of the basic galaxy properties we adopted for the dynamical modeling. The input data required for the construction of the dynamical models (Sec. 2.4) are the 3D stellar luminosity density ν and the spatially resolved line-of-sight velocity distributions (LOSVDs) of the stars. These need to be obtained from photometric and spectroscopic data sets respectively. In the following section we present the data sets we used and outline the procedure to retrieve the luminosity density ν (Sec. 2.1) and the LOSVDs (Sec. 2.2). In Sec. 2.3 and Sec. 2.4 we briefly describe the (stellar population and dynamical modeling) techniques we employed to obtain the information about the structure of the dwarf ellipticals.

In App. E we provide a summary of the morphological peculiarities of each galaxy. To ease the identification of individual galaxies we keep the same color-coding scheme throughout the paper (see Tab. 1, Fig. 2, Fig. 15, or Fig. 23 for the connection between color and Virgo Cluster Catalog (VCC) labels). The basic properties we derive throughout this work are tabulated in Tab. 2.

2.1. Photometry & Deprojection

To obtain a model for the deprojected luminosity distribution ν we draw on publicly available *HST* data. For the majority of the galaxies we use photometry observed with ACS/WCS in the F850LP and F475W filters calibrated in the AB system (cf. Sirianni et al. 2005). One galaxy, VCC 308, has no recorded ACS data, and instead we used WFPC2 observations available in F814W and F555W filters. In the following we will abbreviate these two filters with their very similar counterparts in the Sloan Digital Sky Survey (SDSS) filter system: The g - and z -band respectively. The *HST* data are not always as deep as one may wish, resulting in deprojections

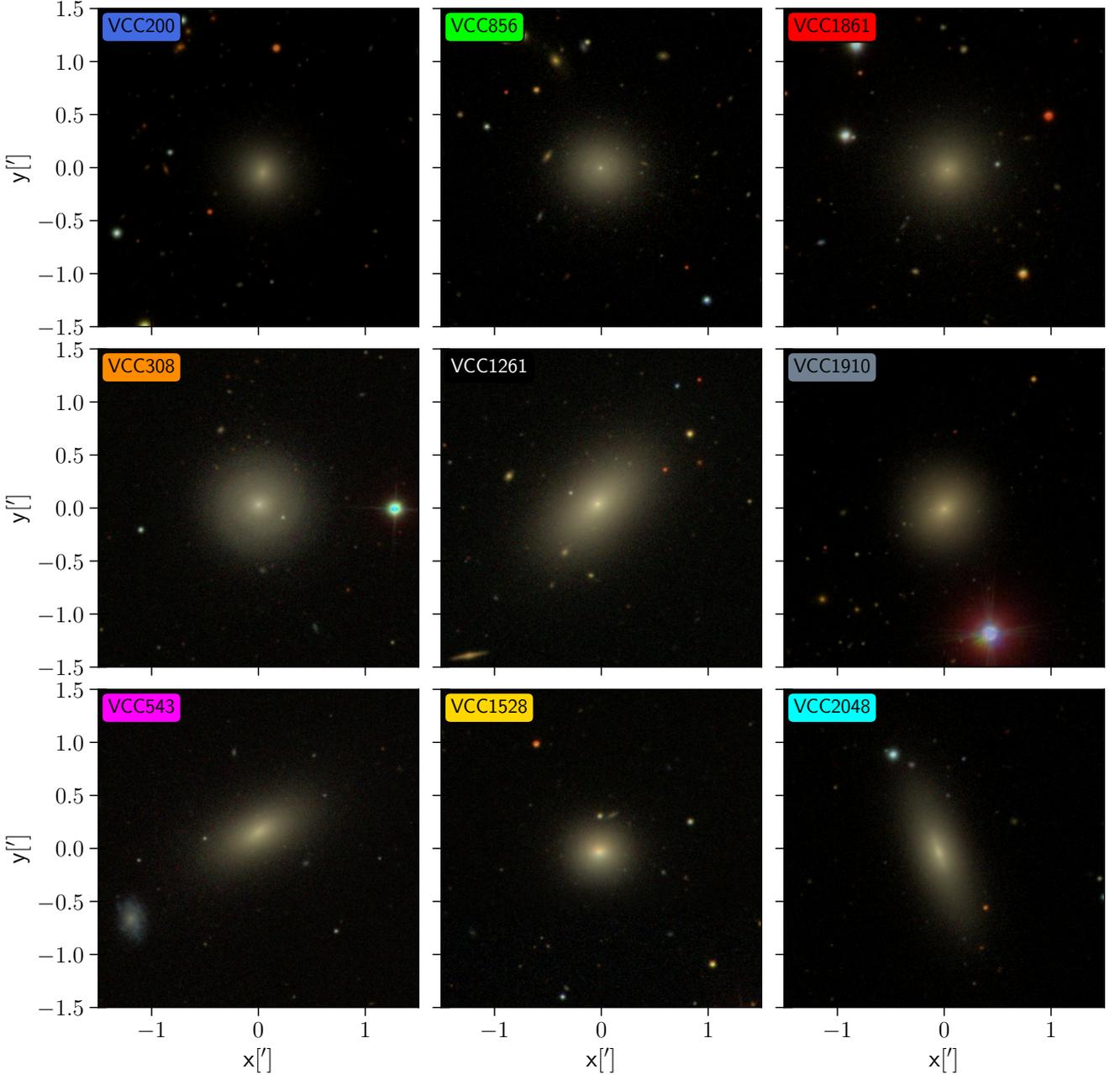


Figure 1. A imaging preview of the dE sample investigated in this study. Shown is the g, r, i image for each galaxy based on publicly available SDSS data. We obtained the images with the help of *Aladin* (Boch & Fernique 2014).

that must be extrapolated to larger radii to cover the full radial extent of a typical orbit model. Therefore, in the case of VCC 1261 and VCC 2048, we chose to re-utilize the more extended photometry of Kormendy et al. (2009) and Kormendy & Bender (2012) instead of the ACS data from the archives. These isophotes are calibrated in the V -band and were extracted using a combination of ACS and SDSS images.

Since the field of view (FoV) of ACS is significantly larger than the size of the dEs, we were able to estimate the background sky value within the science image by calculating the median count in several sky boxes around the galaxies. A good sky estimation is essential for the recovery of the dark matter halos of the dEs. Since the outskirts of the dwarfs are barely brighter than the typical sky background, a bad subtraction may significantly distort the shape and slope of the resulting luminosity

Table 1. Basic properties of the galaxy sample. From left to right: The Virgo Cluster Catalog (VCC) identification number (Binggeli et al. 1985). The color identification used throughout this article. The morphological sub-classification following Lisker et al. (2007) (bc: blue center, di: disk feature, N: classical dE with nucleus, nN: classical dE without nucleus). The distance adopted from Blakeslee et al. (2009), if there are no distances available (other than from redshifts) we use the average Virgo Cluster distance of 16.5 Mpc. The heliocentric velocity cz we measured from the spectra. The projected distance to M87 (assuming $d = 16.5$ Mpc). The stellar mass we obtained. The photometric band we use for the modeling of the light distribution. Effective radius (in the z -band), Sersic-index, and position angle all adopted from (cf. Ferrarese et al. 2006) who investigated all galaxies in our sample except for VC 308 (for which we consulted Hyperleda (Makarov et al. 2014) instead). VCC 308 and VCC 1861 appear almost perfectly round, therefore a robust PA is not given.

VCC ID	Color	Classification	d [Mpc]	cz [km/s]	$\Delta r_{\text{sky, M87}}$ [Mpc]	$\log_{10}(M_*/M_\odot)$	Band	r_e ["]	n_s	PA [°]
VCC 200		dE(N)	18.3	16	1.018	8.85	F850LP	13.12	1.933	-7
VCC 308		dE(bc)	16.5	1527	1.556	8.88	F814W	11.40	1.340	-
VCC 543		dE(nN)	15.8	962	0.905	9.37	F850LP	18.29	1.716	-53
VCC 856		dE(di)	16.9	1016	0.755	8.88	F850LP	16.70	1.317	80
VCC 1261		dE(N)	18.2	1861	0.466	9.44	V	20.13	2.135	-47
VCC 1528		dE(nN)	16.3	1614	0.343	9.01	F850LP	9.88	2.101	84
VCC 1861		dE(N)	16.1	636	0.795	8.88	F850LP	18.24	1.593	-
VCC 1910		dE(di)	16.0	241	0.818	8.85	F850LP	12.01	1.564	-49
VCC 2048		dE(di)	16.5	1096	1.320	9.38	V	12.64	1.973	19

Table 2. Some of the basic properties derived in this study (see text for details). Dark matter and super massive black hole properties will be tabulated in VW–II. From left to right: The VCC identification of the galaxy. The stellar population age (brackets indicate that value is the *average* of the two population ages we derived at around $r = 2.5''$ and $r = 7.5''$, respectively). Similarly the *average* metallicity $\langle [Z/H] \rangle$, abundance ratio $\langle [Mg/Fe] \rangle$ and IMF parameter $\alpha_{\text{IMF}} = \Upsilon_{\text{dyn}}/\Upsilon_{\text{Group}}$. The population values for VCC 1910 are highlighted with a star because they may be compromised (cf. Sec. 2.3). The population parameters $\lambda_{e/2}$ and λ_e within half and one stellar effective radius respectively. Values with a star had to be extrapolated beyond the last kinematic data point. The total stellar angular momentum j_* . The inclination i derived via dynamical modeling. The ellipticity $\epsilon_{e/2}$ of the isophotes within $r_e/2$. The average cylindrical anisotropy parameters $\bar{\beta}_z$ and $\bar{\gamma}$.

VCC ID	$\langle \text{Age} [\text{Gyr}] \rangle$	$\langle [Z/H] \rangle$	$\langle [Mg/Fe] \rangle$	$\langle \alpha_{\text{IMF}} \rangle$	$\lambda_{e/2}$	λ_e	$\log(j_* [\text{kpc km s}^{-1}])$	i [°]	$\epsilon_{e/2}$	$\bar{\beta}_z$	$\bar{\gamma}$
VCC 200	11.6	-0.67	0.23	0.485	0.208	0.211	0.957	90	0.189	-0.038	-0.193
VCC 308	4.3	-0.39	0.11	0.775	0.222	0.233	1.014	46	0.138	0.019	0.060
VCC 543	7.3	-0.39	0.24	1.319	0.375	0.352*	1.488	67	0.456	0.342	-0.005
VCC 856	8.8	-0.55	0.25	0.574	0.330	0.301*	1.719	32	0.093	0.333	-0.156
VCC 1261	7.5	-0.37	0.15	0.767	0.046	0.041*	1.085	65	0.275	0.410	-0.075
VCC 1528	6.0	-0.24	0.16	0.897	0.028	0.027	0.307	90	0.229	0.030	0.013
VCC 1861	9.6	-0.25	0.16	0.492	0.109	0.113	1.176	44	0.053	0.020	0.095
VCC 1910	3.6*	0.25*	0.12*	1.84*	0.126	0.101	1.020	90	0.199	0.135	0.012
VCC 2048	3.6	-0.20	0.22	2.80	0.256	0.225*	1.366	70	0.554	0.372	-0.401

distribution ν . In that case dark mass components of the model will have to compensate for any missing or excessive luminous mass in order to reproduce the correct combined gravitational potential. Furthermore an accurate recovery of ν is important because it serves as a boundary constraint for the stellar mass density in the dynamical models (Sec. 2.4).

After subtracting the sky and masking problematic regions we performed elliptical isophote fits for each

galaxy¹. Where available we compared our isophotes with previous measurements (Ferrarese et al. 2006) and confirmed that we were reliably reproducing them.

The dynamical models we employ are axisymmetric and therefore they require axisymmetric luminosity distributions that can reproduce the observed photome-

¹ We implemented the isophote fitting using python, the code made use of astropy routines (Astropy Collaboration et al. 2022). While the isophote models allowed deviations from an ellipse, the a_4 , a_6 profiles were often quite noisy but consistent with zero for the majority of the dEs.

try. We deprojected each galaxy using the Metropolis-Algorithm of Magorrian (1999), which allows us to explore the full range of axisymmetric solutions for any given inclination i . For our sample, external inclination estimates (e.g. using gas disks) are not possible and we need to probe different inclinations in the dynamical modeling itself. In practice one is limited to testing a small number of deprojections otherwise the modeling quickly becomes computationally unfeasible. To make the modeling as efficient as possible we decided to choose the inclinations i such that the corresponding intrinsic flattening $q(i) = b/a$ of the resulting luminosity densities are (roughly) spaced linearly in the stellar axis ratio q . This maximizes the *intrinsic* physical differences of the densities rather than the absolute difference in the viewing angle itself. The strategy to sample the intrinsic shapes instead of viewing angles is not novel and has found frequent use in triaxial modeling (van den Bosch & de Zeeuw 2010; Walsh et al. 2012; Zhu et al. 2018; Jin et al. 2019; Poci et al. 2019). For a thorough discussion of this sampling choice see van den Bosch & van de Ven (2009); Quenneville et al. (2022). We also took advantage of the fact that the range of possible viewing angles is a priori limited by the observed photometry (e.g. de Nicola et al. 2020). In the axisymmetric case this means we only need to sample the inclination starting from some minimum allowed inclination i_{\min} (consistent with a flat disk).

It is well established that axisymmetric deprojections generally do not have a unique solution for a given viewing angle unless the inclination is exactly edge-on (cf. Rybicki 1987) meaning different boxy and disk deprojections can result in the same projected photometry (cf. Gerhard & Binney 1996). We decided to probe *one* deprojection per inclination as the dynamical differences between boxy and disk deprojections for our dwarf galaxies are typically much smaller than they are between deprojections at different inclinations. The deprojections were chosen to be sufficiently smooth and close to elliptical while still fitting the data adequately.

Table 1 shows the filter band we worked with to obtain the deprojections for the dynamical modeling. Throughout this paper we state all mass-to-light ratios, luminosities, etc. in this corresponding band.

2.2. Spectroscopy & LOSVD extraction

We took the spectra of the dwarf galaxies with the integral-field-unit (IFU) spectrograph VIRUS-W mounted to the Harlan J. Smith telescope at the McDonald Observatory. Table 3 summarises the various observation runs. The 267 fibres of VIRUS-W, which each cover a $3.2''$ diameter on the sky, combine to a field

of view (FoV) of $105'' \times 55''$. Since the velocity dispersion of dEs is generally very low, the instrument was operated in its high resolution mode achieving a spectral resolution of $R = 7900$ to 9000 (or 14 km s^{-1} to 16 km s^{-1}) within the optical wavelength range from 4850\AA to 5475\AA (Fabricius et al. 2008, 2012). In App. B we demonstrate that such a high resolution is essential to study the low dispersions of the dEs. Previous studies with lower resolutions $R < 5000$ are often biased towards too high dispersions (in some cases by up to 50%).

For each galaxy we obtained multiple dithered exposures with a median seeing of $2.0''$. The offsets between the different exposures in the dither pattern range from $1.7''$ to $3.7''$. The small size of the dEs compared to the FoV allowed us to perform the sky correction using the science frames themselves. To reduce the data we used a pipeline based on the ‘Cure’ and ‘Fitstools’ package which was designed for the HETDEX project (Gössl & Riffeser 2002; Hill et al. 2004; Goessl et al. 2006; Hill et al. 2021).

From the dithered exposures a 3D data cube was generated, forming a regular grid of $1.6''$ pseudo spaxels. We binned the pixels further using Voronoi tessellation (Cappellari & Copin 2003) with the goal of obtaining spatial bins with an approximately uniform and sufficiently large signal-to-noise ratio S/N . In the case of our dwarf galaxies this proves to be difficult because they are relatively small compared to the size of the fibres which leads to a relatively large bin-to-bin S/N gradient even after the Voronoi Binning. We found that bins with a $S/N \lesssim 15$ rarely provide stable and reliant LOSVD shapes and exclude them from the further analysis. The mean S/N of the useful Voronoi bins is listed in Table 3. The stellar kinematics cover one effective radius r_e , the sample median FoV size being $1.02 r_e$ (see also Fig. 2).

We derived LOSVDs from the binned spectra using the spectral-fitting code WINGFIT (Thomas et al. in prep.) which allows both a fully non-parametric and also a Gauss-Hermite description of the LOSVDs (see below). The LOSVDs are extracted from each of the Voronoi binned spectra by convolving a model of the LOSVD with a weighted sum of stellar templates. To match the high-resolution of VIRUS-W we used the ELODIE library of template stars in its low resolution version as they cover wavelengths from 3900\AA to 6800\AA with a resolution of $R = 10000$ (Prugniel & Soubiran 2001; Prugniel et al. 2007b). The library consists of spectra including stellar atmosphere parameters with temperatures $T \in [3000K, 60000K]$, surface gravity $\log(g) \in [-0.3, 5.9]$, and $[\text{Fe}/\text{H}] \in [-3.2, 1.4]$. Individual abundance ratios are not resolved in the library

Table 3. Basic Information about IFU observation runs which we attained with the Harlan J. Smith telescope at the McDonald Observatory (Sec. 2.2). From left to right: The Virgo Cluster Catalog (VCC) identification. The month and year of the observation runs. The total exposure time. The total number of Voronoi-binned spectra that survived our restrictions (cf. Sec. 2.2) and were dynamically modelled. The mean Signal-to-noise ratio S/N of said Voronoi bins.

VCC ID	Runs (Month and year)	Total exposure time[min]	N_{bin}	S/N_{bin}
VCC 200	Apr 2016, May 2017	1620	18	21.2 ± 7.1
VCC 308	Feb 2018, Mar 2018, Apr 2018	990	15	17.8 ± 3.3
VCC 543	Feb 2018	1410	16	27.0 ± 6.8
VCC 856	Feb 2014, Mar 2014	1050	18	21.9 ± 4.8
VCC 1261	Feb 2013, Apr 2013	630	45	33.2 ± 7.6
VCC 1528	Mar 2018	450	16	19.6 ± 5.1
VCC 1861	Mar 2014, Apr 2016	1050	29	22.0 ± 6.0
VCC 1910	Jun 2012	510	26	25.0 ± 7.0
VCC 2048	Apr 2012, Feb 2013	1050	36	42.9 ± 9.0

but expected to match the solar neighbourhood patterns. This may result in template mismatch if both metallicity and e.g. $[\alpha/\text{Fe}]$ differ significantly from solar (Prugniel et al. 2007a). Fortunately dEs are expected to have abundance patterns similar to those of LTGs (Şen et al. 2018). Reassuringly, even the most *non*-solar stellar population in our sample, e.g. the central population of VCC 200 with $[\alpha/\text{Fe}] = 0.29$ can still be fitted well using the ELODIE library (see for example Fig. 21). To minimize potential effects of template mismatch on the LOSVD recovery we employed some of the strategies discussed by Mehrgan et al. (2023), which includes optimizing the number of polynomial orders for the continuum fit of the spectra and a preselection of template spectra. For all dwarf galaxies we only fit absorption features as we did not find any significant emission lines within the VIRUS-W spectral range. We inspected each spectrum individually and masked noise contaminated regions near the edges of the spectral wavelength ranges and near cosmic ray hits and sky lines that survived the data reduction pipeline. Fig. 21 in App. A shows a typical Voronoi-binned spectrum obtained with VIRUS-W and the corresponding fit we obtained with the convolved model.

The aforementioned model of the LOSVD can be characterized in two different forms. The first option is a description using a suitable parametric function, typically a Gauss–Hermite series truncated at some highest non-zero order n_{GH} (cf. van der Marel & Franx 1993). The second option is using a more general non-parametric model with some smoothing penalty (e.g. a second derivatives penalty). However, both cases suffer from a similar problem: For Gauss–Hermite models it is unclear for which order n_{GH} the series should be truncated. While for the non-parametric models it is unclear how strong the smoothing penalty strength α_S should be. Both issues are essentially the same issue

of finding the right balance between overfitting and underfitting the spectra: If n_{GH} is too large (the penalty strength α_S too small) the model LOSVD can become arbitrarily complicated, and the model will be overfitting the noise in the spectra. In contrast, if n_{GH} is too small (the smoothing penalty α_S too strong) the LOSVDs will be overly smooth and will not emulate the, perhaps more complex, structure of the underlying stellar motions. Often this problem is passed over by choosing a (hopefully) suitable n_{GH} or α_S or calibrating these factors with mock simulations (e.g. van der Marel & Franx 1993; Ocvirk et al. 2006; Liepold et al. 2020; Falcón-Barroso & Martig 2021). To avoid this we derived a generalized information criterion AIC_p (cf. Thomas & Lipka 2022) which provides a more systematic approach by minimizing the statistical information loss. Defined as $\text{AIC}_p = \chi^2 + 2m_{\text{eff}}$ it penalizes the goodness-of-fit χ^2 a model achieves with its effective flexibility m_{eff} (see Lipka & Thomas 2021; Thomas & Lipka 2022). One finds the optimum amount of model complexity by comparing the AIC_p of model with different degrees of smoothing α_S (or n_{GH}) with one another. The AIC_p criterion is very general and allows us to compare the performance of a Gauss–Hermite model directly with non-parametric model descriptions, thus, allowing us decide what the best way to describe the LOSVD for each individual spectrum.

With this tool in hand we decided to employ the following strategy: For every Voronoi binned spectrum we perform the kinematic extraction using both a Gauss–Hermite parametrization and a non-parametric model with various values of n_{GH} and α_S respectively. We then calculate and compare the corresponding AIC_p for each combination and choose the model that achieves the minimum AIC_p to represent the LOSVD in the given Voronoi Bin (cf. Thomas & Lipka 2022). To provide a coherent input for the subsequent dynamical modeling

we describe all models in the same velocity-binned form no matter if the optimum model is a Gauss–Hermite series or a non-parametric description. We estimated the errors in the velocity bins for both model types by re-fitting 100 Monte-Carlo realizations based on the flux noise (assumed to be gaussian) in each wavelength bin of the spectra. The errors in each velocity bin are quantified as the standard deviation of these realizations. However, since a Gauss–Hermite expansion (by construction) suppresses the error in the high-velocity wings (cf. Fig. 22) we decided on the conservative approach to always adopt the larger error estimate of the two (Gauss–Hermite error or non-parametric error) for each of the velocity bins.

When comparing the non-parametric LOSVDs with the corresponding ones derived from the optimized Gauss–Hermite expansion in the same Voronoi Bin we find that both are fairly consistent with each other, thus, strengthening our confidence in the derived kinematics. A comparison of two typical LOSVDs derived from the same spectrum once by fitting a Gauss–Hermite series and once by fitting a non-parametric description is shown in Fig. 22. Only in the high-velocity tails, where (depending on the maximum order n_{GH} of the Gauss–Hermite) the LOSVD signal tends to be suppressed, we find minor differences between the models. This was already recognized by Mehrgan et al. (2019). Broadly speaking we find that in Voronoi bins with a relatively high S/N the spectra are often preferred to be modelled non-parametrically, while in bins with lower S/N the AIC_p criterion tends to favor Gauss–Hermite expansions with a smaller maximum order n_{GH} (sometimes even gaussian, i.e. $n_{\text{GH}} = 2$). This is to be expected as AIC_p is designed to prevent overfitting noise, i.e. if the noise is more dominant than any underlying LOSVD substructure it is preferable to obtain a smoother description of the main peak of the LOSVD than it is to fit the noise. In Figs. 2 and 3 we show the Gauss–Hermite representations of the final LOSVD data if one approximates each LOSVDs a posteriori as a Gauss–Hermite expansion² with a fixed $n_{\text{GH}} = 4$. This allows for an easier comparison to literature values (App. B) which usually only consist of the first two moments: the rotation v and velocity dispersion σ . We note that the full non-parametric LOSVDs that we used for the dynamical modeling might differ slightly since moments higher than h_4 are not displayed. Similarly, Figure 3 shows the Gauss–Hermite description of the LOSVD data when

compared to the corresponding kinematics of the best dynamical model we found for each galaxy with the technique as described in Sec. 4. Within the errors, the dynamical models we constructed emulate the observations well.

2.3. Single Stellar Population modeling

The stellar population analysis generally requires a higher S/N spectrum than the kinematic extraction we used to obtain the LOSVDs (Sec. 2.2). Since the VIRUS-W Voronoi bins we used for the kinematics have fairly low S/N (Tab. 3) we decided to re-bin the spectra into two radial annuli centered around $r = 2.5''$ and $r = 7.5''$. This allows reliable global estimates of the population quantities while still being able to notice a possible radial variation in the stellar population if present. We can also compare the mass-to-light ratio gradients obtained from the population analysis with the mass-to-light ratio gradients derived from the dynamical models and identify a possible radial variation of the IMF (e.g. Mehrgan et al. 2024; Parikh et al. 2024).

The majority of the Virgo dEs is known to host small blue nuclei with core sizes of $0.2'' - 0.4''$ (Ferrarese et al. 2006; Hamraz et al. 2019). Paudel et al. (2011) separated the nuclei from the galaxy’s main body and found that the nuclei are often composed of a much younger, more metal-rich stellar population compared to the main body of the galaxy. For their nucleated dEs (including VCC 308, VCC 856, VCC 1261, VCC 1861) they found the contribution of the nucleus to the total light within the centre $r \leq 0.375''$ usually does not exceed 50%. Even though we mask the central $0.8''$ before adding the spectra (i.e. the nuclei are not part of the $r = 2.5''$ annulus) the large VIRUS-W fibres with a diameter of $3.2''$ and the atmospheric seeing of $\sim 2.0''$ could lead to a partial contamination of the inner annulus with light from the nuclei. We estimate an upper bound for this contamination by comparing the *total* luminosity of the blue nucleus to the luminosity of the galaxy main body within the circular area covered by the $2.5''$ aperture (incl. the masked centre). The luminosities are obtained by fitting a two-component model to the *HST* photometry, where we assume a King model for the nucleus and a Sersic model for the galaxy main body. For our dE sample we find a median light contribution of 1.3% in the z -band and 1.6% in the g -band with a maximum contribution of 10% for VCC 856. This implies that, even if we were not masking the central $0.8''$, the contamination from the nucleus to the galaxy light within $r = 2.5''$ would be small to insignificant. Hence, our annuli are not affected much by the central nuclei but are well suited to

² We provide a supplementary table for each dE containing these Gauss–Hermite moments, their errors, and pixel location.

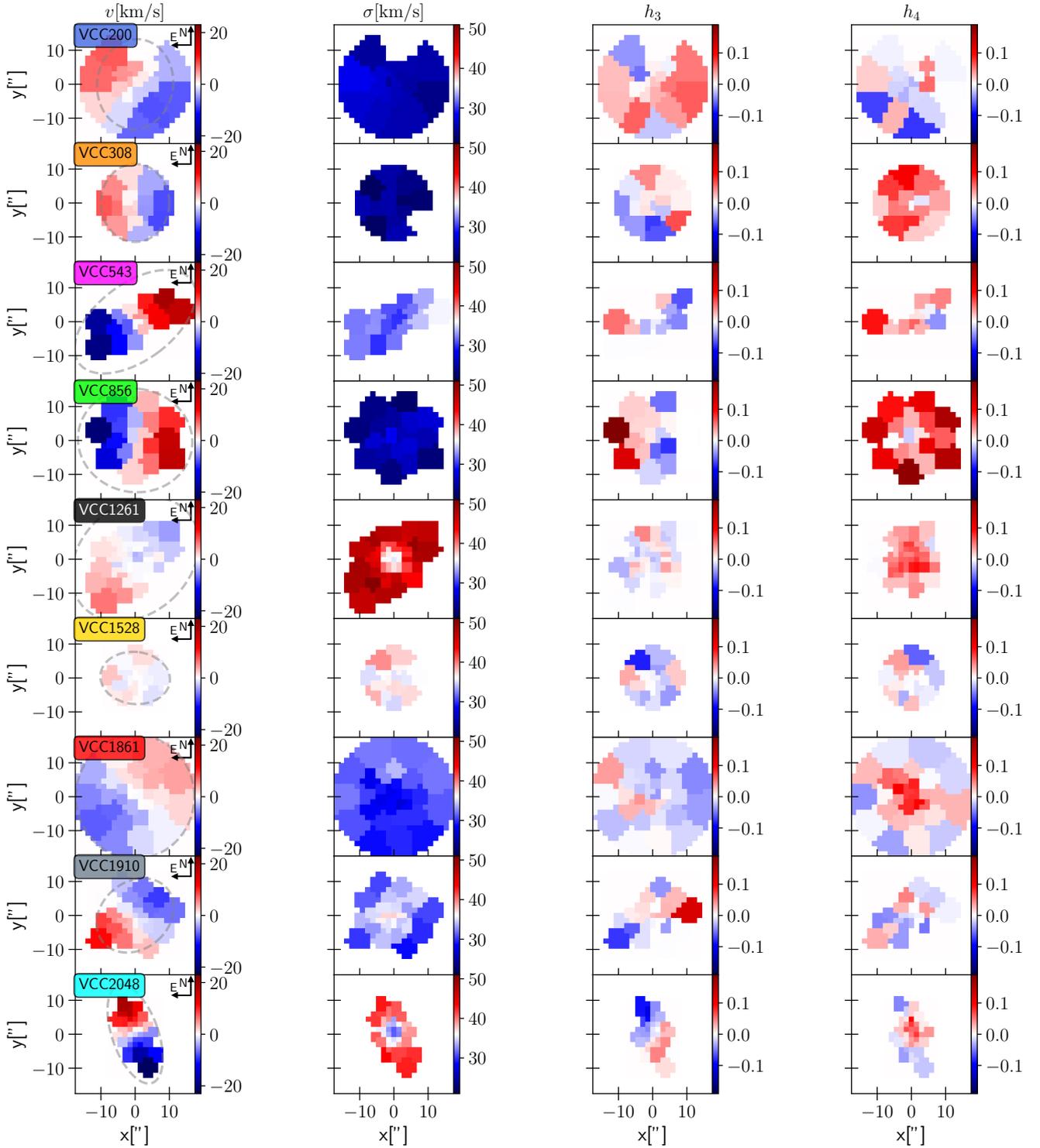


Figure 2. The kinematic maps of the LOSVDs approximated by Gauss–Hermite expansion with fixed $n_{\text{GH}} = 4$, i.e. this does not show potentially higher order deviations. In a few bins the recovered LOSVDs are essentially Gaussian because the optimum n_{GH} was 2. This makes them appear white in the $h_{3,4}$ -maps (the northern most bin of VCC 200 was excluded due to its poor spectrum). We use a *common* spatial and color scale (see colorbars on the right sides of the panels) to enable an easy comparison between *different* sample galaxies. Spatial variations within a given galaxy are more discernible in the radial profiles (Fig. 3). In the velocity panels we indicate the North-East directions and the photometric $1.0 r_e$ aperture (*dashed ellipse*).

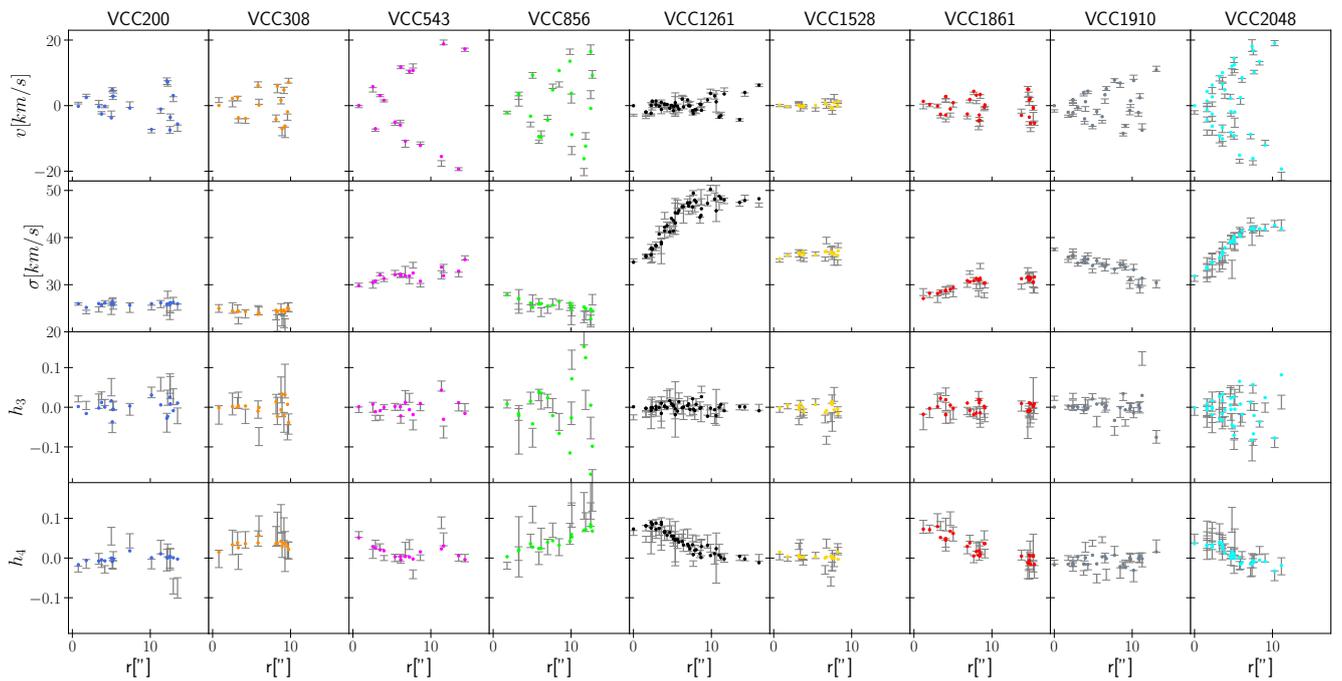


Figure 3. The radial profiles of the Gauss–Hermite representations (up to h_4) with 1σ -errors as derived from the spectroscopic data (*gray*) and the corresponding fit of the best dynamical orbit model we found for each galaxy (*colored*). The 1σ -errors are quantified by fitting Gauss–Hermite to 100 Monte-Carlo realizations on the velocity binned LOSVDs. The models fit the observed kinematics quite well considering deviations from axisymmetry in the data cannot be emulated by axisymmetric models.

focus on detecting gradients within the main body of galaxy.

The SSP models were derived from the absorption line indices by modeling them with the index models of Thomas et al. (2011a). Within the VIRUS-W spectral range we model 4 notable Lick indices: $H\beta$, Mgb , $Fe5270$, $Fe5335$ to obtain the light-weighted age, metallicity $[Z/H]$, and the abundance ratio $[Mg/Fe]$ of the stellar population. The models were probed by sampling a grid spanned by these 3 parameters and their best fit values were derived by interpolating the grid in χ^2 . The population models are resolution corrected to match the data and the uncertainties are based on 100 Monte-Carlos simulations assuming gaussian noise for the flux in each wavelength bin (more details in Parikh et al. 2018). The best fit parameters and their relation to each other are shown in Fig. 4. We find that old dEs are metal-poor while younger ones have nearly solar-like metallicity. Conversely the $[Mg/Fe]$ ratios are slightly over-abundant and show no clear correlation with the metallicity.

VCC 1910 has an exceptionally high, super-solar metallicity (see Fig. 4). We suspect that the SSP analysis for this galaxy is compromised (especially for the inner spectrum) because the most important age-sensitive Lick index within our spectral range ($H\beta$) lies just at the edge of the usable wavelength region which makes a robust derivation of the Lick continuum difficult. While we still show the results for VCC 1910 for the remainder of this work, one should take the SSP results for this galaxy with a grain of salt. We refer to the literature SSP values for VCC 1910 (App. B).

2.4. The Dynamical modeling technique

The dynamical modeling code we employ is the current version of the axisymmetric orbit superposition code of Thomas et al. (2004), which is continuously updated and is based on the Schwarzschild orbit superposition approach (Schwarzschild 1979). Schwarzschild modeling is a versatile numerical approach and can in principal be applied to any collisionless system. For a detailed description of the modeling we refer to Thomas et al. (2004).

To briefly summarize: The technique can be broken down into a few steps. A manifold of trial density mass models ρ is constructed and for each a corresponding gravitational potential Φ is calculated. In each potential a set of representative orbits is integrated which densely sample the phase-space compatible with said potential Φ . Each orbit is assigned an occupation weight and all orbits are superimposed and the model properties are determined by the weighted superposition. The best

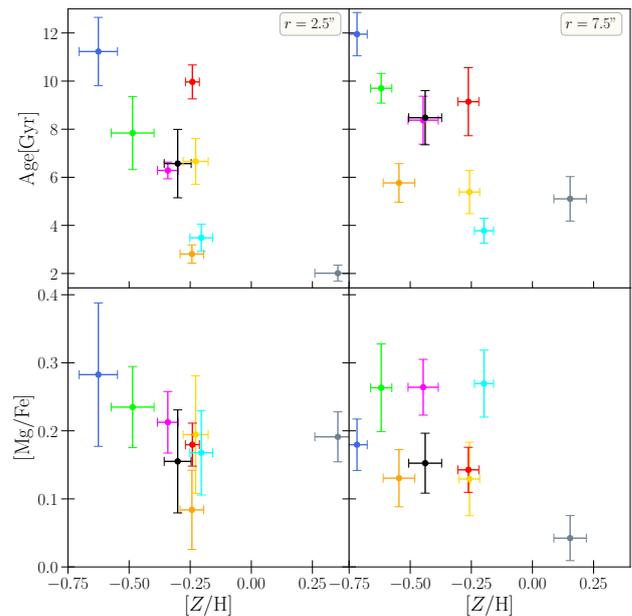


Figure 4. Correlations of the SSP parameters for the dE sample. Shown are the results for the spectra in the two apertures centered around $r = 2.5''$ (left) and $r = 7.5''$ (right). Top: Age vs metallicity. Bottom: $[Mg/Fe]$ vs metallicity.

set of weights is found by fitting a model’s observables to the corresponding observations, which in our case are the spatially resolved, non-parametric LOSVDs of the stars. The number of orbits, and thus the number of weights, is typically larger than the number of data constraints. Therefore a regularization term is required that prevents overfitting. In Lipka & Thomas (2021); Thomas & Lipka (2022) we presented and applied a novel data-driven approach to optimize the degree by which the models should be regularized to avoid both, over- and underfitting. This is the same AIC_p approach we already utilized in the kinematic fitting procedure (cf. Sec. 2.2). It can generally be used to optimize penalty strengths in any penalized model fitting process. After fitting the weights for each of the trial mass models, the quality of each model can be compared by the AIC_p it achieved: The single model which achieves the smallest AIC_p is then deemed to be the best representation of the galaxy that’s under investigation. In Lipka & Thomas (2021) we demonstrated and discussed why the AIC_p selection is more appropriate and unbiased in finding the best model out of all the probed trial models when compared to a simple χ^2 selection. This is because AIC_p accounts for the intrinsically varying model flexibility across different orbit models whereas χ^2 only evaluates how well the data was fitted, making it prone to overfitting.

A reasonable approach is to describe the mass model with a mass distribution consisting of 3 components:

$$\rho(\mathbf{r}) = \Upsilon_* \cdot \nu + \rho_{\text{DM}} + M_\bullet \cdot \delta(\mathbf{r}) \quad (1)$$

Where M_\bullet is a central supermassive central black hole (SMBH), ρ_{DM} is a description of the dark matter halo (usually given in a parametric form), and Υ_* is a stellar mass-to-light ratio which together with the 3D luminosity distribution ν describes the distribution of the baryonic matter. The luminosity distribution ν functions as a boundary condition for the models and is obtained from the photometry (Sec. 2.1).

For the construction of the trial mass models we follow this standard 3-component mass description. We find that to reliably decompose different mass components, a flexible and generic description of the *dark matter* component is crucial. Otherwise the recovery of the other mass components (stars and SMBH) may be negatively affected as they may try and compensate for the inflexibility of the halo model. In Lipka & Thomas (in prep.), in the following abbreviated as LT, we will lay out our arguments for why we believe that a sufficiently unbiased, yet still computationally efficient, decomposition is achieved if we parametrize the dark matter with a (flattened) Zhao-profile (Zhao 1996) where the transition width parameter is fixed. This means the halo component of our mass models is described by the following elliptical profile:

$$\rho_{\text{DM}}(m, \theta) = \frac{\rho_0}{\left(\frac{m}{r_s}\right)^{\gamma_{\text{in}}} \cdot \left(1 + \frac{m}{r_s}\right)^{\gamma_{\text{out}} - \gamma_{\text{in}}}} \quad (2)$$

where the ellipsoidal radius m and angle θ are the (oblate) ellipsoidal coordinates within the meridional plane. This model has an inner logarithmic density slope γ_{in} and a corresponding outer slope γ_{out} . The scale radius r_s locates the transition between the two slopes while ρ_0 sets the density scale. Together with the halo flattening q_{DM} , this means the halo is described by a 5-parameter halo model, much more than the standard 1- or 2-parameter models that are commonly used. Since the dark matter component is the subject of VW-II, we discuss our halo parameter sampling strategy in more detail in that study.

We do not know a priori by how much the galaxy is inclined, which not only affects the projection to the sky of the entire model but also the photometric boundary constraint ν . Therefore we sample differently inclined orbit models and deprojections (see also Sec. 2.1). In the past it was rather common to model only a single (edge-on) deprojection per galaxy. This practice was reinforced by the question whether it is even possible

to dynamically constrain the inclination of axisymmetric galaxies because there appears to be a significant χ^2 bias favoring edge-on models over less inclined models (see Lipka & Thomas 2021 for an in-depth explanation of the origin of this bias). In Lipka & Thomas (2021) we demonstrated that the viewing angle of a galaxy can be well constrained using dynamical models directly as long as one follows the model selection approach we developed (cf. Lipka & Thomas 2021; Thomas & Lipka 2022) and described above.

Apart from the viewing angles, we also do not know whether the assumption of a constant stellar mass-to-light ratio is a reasonable one. For example the stellar mass-to-light ratio could vary with radius because the IMF (e.g. Mehrgan et al. 2024) or the age and composition of the stellar population could exhibit spatial variations within the regions probed by our dynamical models. Similar to an inflexible dark matter component, an erroneous stellar component would also have effects on the recovery of all other mass components as those will try and compensate for this as much as possible to emulate the overall gravitational potential well. For this reason we allowed for another parameter that increases the flexibility of the *stellar* model component. We allow for a radial variation in the stellar mass-to-light ratio Υ_* . This is implemented by independently sampling an inner and outer mass-to-light ratio Υ_i , Υ_o at two fixed ellipsoidal radii m_i and m_o . Between the two radii the mass-to-light ratio is interpolated log-linearly, while outside of it ($m < m_i$ and $m > m_o$) the mass-to-light ratio is kept constant at Υ_i and Υ_o respectively. The ellipsoidal coordinates on which the mass-to-light ratios are stratified have the *average* intrinsic axisymmetric flattening $q(i)$ inherited from the stellar luminosity deprojection $\nu(i)$ at inclination i . For a more detailed explanation of the gradient implementation see also Mehrgan et al. (2024).

The choice of radii at which one samples the two mass-to-light ratios is non-trivial and several issues have to be considered: Dynamical models have reduced constraining power at scales smaller than the spatial resolution and in the areas not covered by data (see LT). Therefore probing Υ_i at radii smaller than the resolution and Υ_o outside the FoV is dangerous and, in the worst case, biases the interpolation of $\Upsilon_*(r)$ between the two radii. This issue should be accounted for in the modeling of gradients for all types of galaxies. However, in the specific case of our dE sample we want to highlight two further issues that could lead to misleading gradients $\Upsilon_*(r)$ if not treated with care. Many of the dEs in our sample host distinct blue nuclei (cf. App. E and Ferrarese et al. 2006) whereas the extended host galaxies exhibit a $g - z$ color that is almost spatially constant.

This suggests a distinct, younger stellar population in the nucleus embedded in a relatively homogeneous older population (Paudel et al. 2011). If one were to sample Υ_i within said nucleus, it could distort the results for the entire galaxy since a log-linear interpolation over the entire radial range would not be a good approximation of the locally much more concentrated young central population. This could become a significant problem if the nuclei are resolved (or just about resolved). Fortunately, in our case the typical size of the blue nuclei is much smaller than the fibre size of our observations and the nuclei only contribute a small portion of light to the central Voronoi bins (see discussion in Sec. 2.3).

Therefore the mass-to-light ratio variation $\Upsilon_*(r)$ we derive should only reflect a (possible) stellar population variation within the larger galaxy envelope itself. The second issue we had to consider when modeling the stellar gradients of the dEs is their relatively faint surface brightness which reaches the level of the sky background for $r \gtrsim 2 r_e$. A small systematic mis-estimate of the sky level will systematically increase/decrease the slope of the measured surface brightness as the relative contribution of the sky to the total light increases with radius. In turn, this affects the luminosity deprojection and the associated dynamical mass-to-light ratios. To be confident in the level of sky that we determined from *HST* data (Sec. 2.1), we carefully checked how well the reduced surface brightness profiles agree with the independently obtained profiles by Ferrarese et al. (2006), and avoided probing Υ_o at very large radii. Considering the above arguments we decided to sample the mass-to-light ratios of the dEs at $m_i = 1.5''$ (~ 0.12 kpc) and m_o between $9''$ to $14''$ (~ 0.7 kpc – 1.2 kpc) depending on the FoV size. That is at about the size of the spatial resolution limit and near the edge of the FoV, i.e. we allow the mass-to-light ratio to vary out to approximately 1 effective radius. This choice avoids the less constrained regions at very small and large radii, mitigates possible distortions by the nuclei or sky, and at the same time still allows for a stellar variation over the entire radial range that we believe is well constrained by the data.

For each galaxy we calculated $10^3 - 10^4$ models on the 9-Dimensional grid of candidate mass models that is spanned by all probed parameters ($\Upsilon_i, \Upsilon_o, i, M_\bullet, \rho_0, r_s, \gamma_{\text{in}}, \gamma_{\text{out}}, \text{qDM}$) which determine the total mass distribution of the model. Since we sample about 5-20 values of each of these parameters, this huge grid can only be probed partially. Therefore we searched the grid efficiently by using the Nonlinear Optimisation by Mesh Adaptive Direct search NOMAD (Audet & Dennis 2006; Le Digabel 2011). We conducted multiple *independent* NOMAD iterations to avoid biases due to the search

algorithm (see LT). We stopped the model calculation when the best AIC_p models all had roughly congruent mass distributions within the regions that are well constrained by data, and when AIC_p had essentially converged to a constant value. As we will discuss in LT, further search might change the nominal values of some of the (correlated) halo *parameters* significantly, yet such a change in parameter values would only result in minuscule changes in the mass-distribution/composition due these correlations. Consequently such parameters should be thought of as *nuisance* parameters only employed to describe the mass distribution.

Error estimates in the data (here the LOSVDs) are often imperfect, non-gaussian, and correlated (Houghton et al. 2006), whereas the modeling implementations assume the errors to be independent and gaussian. As long as these issues are statistically unaccounted for, the *absolute* values of χ^2 , and consequently AIC_p , are not meaningful statistically as they do not reflect the correct level of noise. For example, correlated raw data suppress the value of χ^2 if independent gaussian noise is assumed. In such a situation, the often applied χ^2 criteria to gauge the goodness of fit and to compute confidence intervals becomes meaningless.

Ideally one would model independent measurements of the same objects and evaluate the resulting scatter to obtain a realistic measure of the error. We can achieve this by modeling the quadrants of axisymmetric systems independently, which has the benefit of not relying on the perfect accounting of noise patterns and even includes possible *systematic* uncertainties (e.g. deviation from axisymmetry, dust, etc.) in its calculation. While this is our preferred approach, the number of spatial bins N_{bin} which survived our S/N limit of 15 is fairly low for the majority of the dEs (Tab. 3). Therefore we were forced to model the FoV as a whole instead of individual quadrants. This has two side-effects: 1) Deviations from axisymmetry across the different quadrants cannot be emulated by the axisymmetric orbit models. 2) We cannot use the scatter of the 4 independent quadrant modelings to estimate an error for the galaxy properties we infer from the modeling.

Therefore, in the case where one can only model a single measurement with imperfect noise accounting, we need to come up with an alternative noise estimation (see details LT). Even if the absolute differences in AIC_p (or χ^2) between different models is not meaningful in itself, we can assume the model evaluation statistic (i.e. AIC_p or χ^2) is at least consistent and unbiased. In that case the relative ranking of the orbit models can still be used to gauge the significance of the resulting dynamical constraints. For example, if the differences in

terms of AIC_p between the best few N models are low, yet these models differ significantly in their recovered M_\bullet , then the confidence in the M_\bullet measurement is correspondingly bad. In order to not *under*-estimate the errors this requires that enough models are included in the calculation of the scatter. And in order not to *over*-estimate it, enough models have to have been probed globally for AIC_p to have approximately converged.

In practice, we tried to ensure an accurate error estimation by probing so many models that neither AIC_p nor the scatter of the N best models changes significantly anymore. We then implemented the error estimation by calculating all errors from the scatter between the best 25 models we found for each galaxy. Given the number of total models we calculated (see above) this translates to a significance criterion where the best $\sim 0.5 - 1.0\%$ of all probed candidate models or analogously all models with roughly $\Delta AIC_p \lesssim 10$ are considered (see also Fig. 10). In conventional statistical modeling a rule of thumb is that models with $\Delta AIC > 10$ are so unlikely they can be excluded (cf. Burnham & Anderson 2002).

We stress-tested the above dynamical modelling strategy and code on a mock VIRUS-W observation of an N-body dwarf simulation with stellar mass $M_{\text{star}} = 4 \cdot 10^8 M_\odot$. The stress-test is setup such that stellar and dark matter are distributed similarly in large regions of space, which we expect to complicate the dynamical decomposition of the two mass components. This allows us to gauge how well we can recover dynamical stellar mass-to-light ratios, dark matter distributions, black holes, and the anisotropy structure even under particularly bad conditions and whether the model parameters could suffer from degeneracies. The results of the mock test are shown in VW-II where we also present all the dynamically recovered properties of the dEs (here only the stellar component is discussed).

3. PROJECTED KINEMATIC STRUCTURE, STELLAR POPULATIONS AND THEIR LINK TO ENVIRONMENT

Before we discuss the dynamical models and *intrinsic* structure of the dEs we first discuss their *on-sky* kinematic structure and stellar populations, and investigate if and how these stellar properties depend on the local environment the galaxies currently inhabit.

3.1. On-sky kinematic structure

Overall, the kinematic moments of the dEs display a large diversity³ which suggests they took distinct evolutionary paths, or are at least at a different stage of their evolution which among other things could depend on their initial mass and/or past interactions with their environment. We find that many dEs have radially increasing velocity dispersions, while others have essentially flat profiles, only two galaxies exhibit a steadily decreasing σ that peaks in the center. The degree to which the dEs are rotation- or pressure-supported also seems to vary widely from galaxies with $v/\sigma \sim 0$ (e.g. VCC 1261, VCC 1528) to galaxies with a substantial amount of rotation $v/\sigma \sim 0.5 - 0.8$ (e.g. VCC 543, VCC 856, VCC 2048). At least in the dwarfs where rotation is a significant factor and the signal-to-noise is high enough to reliably constrain higher moments, we find that the higher Gauss-Hermite moments of dwarf ellipticals follow the well established $v - h_3$ anti-correlation known from massive ellipticals (Bender et al. 1994). The h_4 profiles we observe show surprising individuality. About half the sample galaxies have a maximum h_4 in their center which steadily drops to 0 with increasing radius. Other galaxies show no clear h_4 trend with radius, i.e. they either scatter stochastically or are consistent with a gaussian LOSVD. Only VCC 856, and to some degree VCC 308, are outliers in having a *rising* h_4 profile which is gaussian in the center. This peculiarity could be linked to the fact that both these galaxies are found to contain weak signatures of face-on spiral arms (cf. App. E), and indeed we will find these two dEs to be the ones that are closest to face-on (Sec. 4).

‘Ordinary’ early-type galaxies⁴ are known to come in two types (Kormendy & Bender 1996), that are often separated into ‘Slow-Rotators’ and ‘Fast-Rotators’ according to their angular momentum parameter λ as defined by Emsellem et al. (2007):

$$\lambda = \frac{\langle r|v| \rangle}{\langle r\sqrt{v^2 + \sigma^2} \rangle} \quad (3)$$

Here the brackets indicate the flux-weighted sum over all spatial bins within a given aperture (usually within half or one stellar effective radius).

³ To highlight this diversity we sum up and discuss the kinematic (and photometric) features of each galaxy individually in App. E.

⁴ In the following the term ‘ordinary’ ETG broadly refers to ETGs with stellar masses $\log_{10}(M_*/M_\odot) \gtrsim 10$. Galaxies classified as dEs (or spheroidals) dominate the population of early types below this mass threshold, while ‘ordinary’ (or classical) ETGs, which differ in their surface brightness distribution (see Kormendy & Bender 2012), are usually found above this mass (there is only a small overlap of these two populations). The ‘ordinary’ ETGs can be sub-classified further (Kormendy & Bender 1996) but this distinction is not considered here.

Galaxies can then be classified by comparing their λ with their apparent ellipticity⁵ ϵ . This allows one to quantify whether the flattening of a galaxy is mostly supported by its angular momentum (i.e. ordered motion) or more by an anisotropy in its stellar velocity dispersion tensor (though anisotropy and rotation can easily go together, cf. Thomas et al. 2009a).

Generally speaking the angular momentum parameter is highest for LTGs (particularly for those on the star-formation main-sequence, Wang et al. 2020) as stars tend to form in rotating gaseous disks. However, even ETGs can have a high angular momentum support and several studies have found a systematic change of λ in the ETG sequence depending on total stellar mass and environment (see below). Since the FoV of our dE sample does not always extend out to $1 r_e$, we decided to evaluate the angular momentum parameter within an aperture of $r_e/2$. Fig. 5 shows $\lambda_{e/2}$ vs $\epsilon_{e/2}$ for our dE sample together with samples of ‘ordinary’ ETGs (Cappellari et al. 2011; Mehrgan et al. 2023) and three dE samples in a similar mass range as ours (Toloba et al. 2015; Scott et al. 2020; Bidaran et al. 2020). For Scott et al. (2020) and Bidaran et al. (2020) the ellipticity was not given in the $r_e/2$ aperture. However, the differences between $\epsilon_{e/2}$ and ϵ_e can be expected to be small (e.g. from Toloba et al. (2015) who show both, the mean difference between the two is ~ 0.019). An alternative to Fig. 5 for the larger $1 r_e$ aperture is shown in App. F. While it relies on using extrapolated λ_e values for some of the dEs when the FoV is too small, none of the following conclusions change.

We estimate the statistical errors of our $\lambda_{e/2}$ values using 100 Monte-Carlo realizations of the velocity and dispersions maps. For each realization the values in each Voronoi bin are perturbed according to their respective measurement error and a new $\lambda_{e/2}$ is calculated from the perturbed values. From the 100 $\lambda_{e/2}$ values we estimate the statistical errors. We find the statistical errors of $\lambda_{e/2}$ are quite small ~ 0.01 (i.e. roughly the size of the markers in Fig. 5). Differences between the λ parameters from different studies are more likely to be driven by systematic effects due to different spatial binning, apertures and spectral resolution (App. B). We are able to estimate the impact of such systematics if we compare our values of those dEs that we have in common with the sample of Toloba et al. (2015). We find that

⁵ The ellipticity is measured within the same aperture as λ . The calculation (or definition) of ϵ is not always entirely consistent across different studies. We follow the definition of Emsellem et al. (2007) which means we are stating the *flux-weighted* ellipticity within the considered aperture.

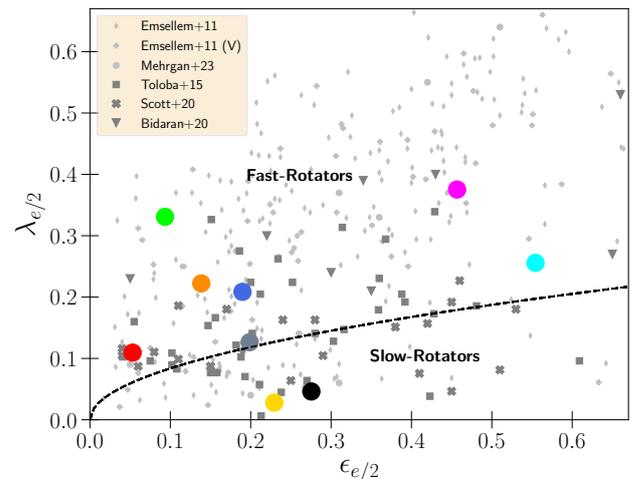


Figure 5. The projected angular momentum parameter λ vs ellipticity $\epsilon_{e/2}$ within *half* the effective radius. The *dashed* curve indicates the dividing line between the Slow- and Fast-Rotator classification (Emsellem et al. 2007). *Colored circles*: Our data. *Gray squares*: The Virgo dEs of Toloba et al. (2015). *Gray X-marks*: The Fornax dEs of Scott et al. (2020). *Gray triangles*: The infalling Virgo dEs of Bidaran et al. (2020). *Light gray crosses*: The massive Virgo cluster ETGs ($M_* \sim 10^{10} - 10^{12} M_\odot$) of Emsellem et al. (2011) which are part of the ATLAS^{3D}-survey (Cappellari et al. 2011). *Light gray diamonds*: The massive ETGs of Emsellem et al. (2011) that are not part of Virgo. *Light gray dots*: The massive ETGs of Mehrgan et al. (2023).

our $\lambda_{e/2}$ are on average 0.077 higher than their 2D extrapolations⁶. Our $\lambda_{e/2}$ are systematically higher than those of Toloba et al. (2015) because of the poorer spectral resolution of their spectra. Low resolution biases the velocity dispersions high (we show this in App. B), which itself results in an underestimation of the angular momentum parameter. While this spectral resolution effect results in a systematic bias we expect the impact of spatial binning to be of a more random nature. And indeed we find no correlation of $\lambda_{e/2}$ for our dEs with the spatial binning (the spatial binning is not homogeneous for all dEs, see Tab. 3). By re-binning the spectra of VCC 2048 into coarser bins and calculating $\lambda_{e/2}$ again we estimate the spatial resolution systematics to be around ± 0.02 . Finally we gauge the systematics in the ellipticity $\epsilon_{e/2}$, again by comparing the dEs in common with Toloba et al. (2015), and find a mean difference of $\Delta\epsilon_{e/2} = -0.037$.

Our dEs are distributed fairly similarly to the other 3 published dE studies, though our sample tends to have

⁶ The values of Toloba et al. (2015) are based on long-slit measurements that were extrapolated to a 2D estimate using a correction factor.

lower ellipticity compared to some of the other studies⁷. However, we note that this comparison is not always straightforward. The λ values of Toloba et al. (2015) are based on long-slit data that were transformed to ‘integrated’ λ values using a 2D-correction factor (see Toloba et al. (2015) for details). The values in Scott et al. (2020) were given only within the larger aperture of $1 r_e$ and some of them are extrapolations using the aperture corrections of van de Sande et al. (2017). We applied the inverse of this correction to estimate $\lambda_{e/2}$ from the λ_e values stated in Scott et al. (2020). These $\lambda_{e/2}$ are shown in Fig. 5. The dEs of (Bidaran et al. 2020) may be selection biased as they are all part of the same, recently accreted, galaxy group and most of them are classified as dE(di). Such dEs can be expected to have a higher angular momentum parameter than the average Virgo dE (see Sec. 3.2). Finally, as suggested above, the poorer spectral resolution of Toloba et al. (2015) has likely biased the $\lambda_{e/2}$ low. In comparison the intermediate resolution of the samples of Scott et al. (2020) with $R = 4500$ and Bidaran et al. (2020) with $R = 3000$ may have been just enough to obtain unbiased estimates if the S/N was high enough (cf. Eftekhari et al. 2022).

Compared to the global picture all the above differences between dE studies are minuscule. Taken together all 4 dE studies (Toloba et al. (2015); Scott et al. (2020); Bidaran et al. (2020) and ours) suggest that the angular momentum parameter of dEs is systematically lower than those of the ‘ordinary’ ETGs. This was first noted by Scott et al. (2020) and even the high spectral resolution power of VIRUS-W does not change this result. Very few dEs have a high angular momentum parameter ($\lambda_{e/2} \gtrsim 0.3$) and the angular momentum parameter is almost independent of the ellipticity. In contrast, the vast majority of ‘ordinary’ ETGs with $M_* > 10^{10} M_\odot$ (excluding the most massive ones) are classified as fast-rotators (Emsellem et al. 2011; Guo et al. 2020; Santucci et al. 2022) and $\lambda_{e/2}$ tends to be more strongly correlated with $\epsilon_{e/2}$.

It is known that ‘ordinary’ ETGs are not homogeneous in λ either. Several studies of the massive ETGs have previously noted a *dichotomy* in the angular momentum parameter as the total stellar mass increases further: The slow-rotator regime is mostly dominated by the most massive ETGs with $M_* \sim 10^{12} M_\odot$ whereas the intermediate mass ETGs ($10^{10} - 10^{11} M_\odot$) tend to be more rotation supported (e.g. Emsellem et al. 2011; Gra-

ham et al. 2018; Santucci et al. 2023). This dichotomy in the ETG sequence has been well established for some time now as several properties change in this regime (e.g. Bender 1988; Bender et al. 1989; Kormendy & Bender 1996; Kormendy 1999; Kormendy & Bender 2013; Dekel & Birnboim 2006; Nelson et al. 2018).

This behaviour of a decreasing λ with increasing stellar mass found for ‘ordinary’ ETGs is opposite to what we observe going from the regime of the dEs to more massive ETGs: an increase of λ with mass. Consequently, if we include the dEs in the ETG mass sequence (i.e. a sequence spanning 5 dex from $\log_{10}(M/M_\odot) = 8$ to 12) we would not find a dichotomy but a ‘trichotomy’ in λ as the total stellar mass changes: The angular momentum parameter is low in the mass regime of the dEs but increases with stellar mass until it reaches a maximum for galaxies with $10^{10} - 10^{11} M_\odot$ at which point the trend reverses and the angular momentum parameter decreases again with the most massive ETGs being slow rotators again. Unfortunately, we are not aware of a *single* study/sample that covers the ETG sequence sufficiently well from $\log_{10}(M/M_\odot) = 8 - 10.5$ to observe the increase in λ with mass directly and we rely on comparison across different galaxy samples with possibly different systematics. See for example Scott et al. 2020 or Spavone et al. 2022 who also observed this trichotomy, or ‘U-shape’, by showing the λ vs stellar mass using a number of different studies. How the relation between stellar mass and λ continues for even lower masses is less clear. The results of de los Reyes et al. (2023) suggest that rotation support decreases even further for galaxies below $\log_{10}(M/M_\odot) \leq 8$.

The λ correlation with mass could suggest that there is a process that suppressed the *total* angular momentum J of the dEs, or conversely, that dEs hide more angular momentum at $r > 1 r_e$ than ‘ordinary’ ETGs do. Fig. 6 shows the estimated stellar specific angular momentum $j = J/M$ (i.e. per mass) vs absolute magnitude for our Virgo dE sample together with other galaxy samples of various morphologies. An alternative version of the Figure vs stellar mass M_* instead of magnitude is shown in App. F.

For LTGs such a diagram is a standard analysis tool, but for ETGs it is much more difficult to determine the total specific angular momentum $j = J/M$ due to the lack of (intrinsic) rotation measurements beyond $\sim 1 - 2 r_e$ and the frequent lack of information about the viewing angles and intrinsic shape. Still it has been done for ‘ordinary’ ETGs (e.g. Romanowsky & Fall 2012; Pulsoni et al. 2023) using careful approximations and information from non-stellar tracers at larger radii. However,

⁷ Considering Lisker et al. (2007) who analysed 413 Virgo dEs (complete in the regime of our dEs) we would expect a blindly selected Virgo dE to have $\epsilon \sim 0.25$ on average.

where dEs (or dwarf ETGs in general) appear in these diagrams has not been addressed yet.

One can obtain a rough estimate of a galaxy’s angular momentum using $j \approx \kappa \cdot v_m \cdot r_e$ where v_m is the velocity around the maximum of the rotation curve, r_e the effective radius, and κ a numerical factor that follows from assumptions about the density and velocity distributions of the stars. κ can have a range of values $\in [1, 4]$ depending on the galaxy morphology and assumptions going into it (cf. Fall 1983; Scorza & Bender 1995; Romanowsky & Fall 2012; Pulsoni et al. 2023).

Following Bender & Nieto (1990); Bender et al. (1991); Scorza & Bender (1995) we analyse *all* ETGs using $\kappa = 1.03$ which is an approximation of j enclosed within $5 r_e$ derived for de Vaucouleurs galaxies (Sersic index $n = 4$). More intricate approximations of j exist that, for example, account for variation of κ with Sersic index (Romanowsky & Fall 2012). However, such 2nd order corrections make it difficult to compare a large number of galaxy samples of different morphologies and our intention here is to investigate the transitional behaviour of j as one goes from ‘ordinary’ ETGs to dEs as ‘fair’ as possible. Relying on different a priori assumptions and corrections for the two galaxy types could distort and complicate this comparison. Still, we stress that an inclusion of such 2nd order corrections would strengthen the evidence for the following conclusion rather than weakening it. For example, dEs generally have lower Sersic indices $n_s \sim 2$ than ‘ordinary’ ETGs (see Tab. 1 or Kormendy et al. 2009) and the correction factor κ is rising monotonically with Sersic index. From the correction factor of Romanowsky & Fall (2012) we expect dEs to have j overestimated by about 50% *relative* to ‘ordinary’ ETGs which have a higher Sersic index ($n_s \sim 4$). Other sources of uncertainty are the poorly known inclinations and the small FoV (i.e. the maximum of the circular velocity curve may not be sampled yet). These are errors expected to be of similar order as the κ correction (Scorza & Bender 1995; Romanowsky & Fall 2012) but we can expect those to affect all ETGs alike (i.e. relative differences remain mostly unaffected).

Nonetheless a comparison across different samples and morphologies is not straight-forward. For the ETG samples shown in Fig. 6 we obtained the maximum of the velocity curve v_m and r_e as stated in these studies⁸, but this does not mean the maximum of velocity curve has actually been probed fully. For example, for our dEs Fig. 3 suggests we are close to the maximum for some

of the dEs but others may still rise beyond the FoV. In Fig. 27 we show one of the, currently most definite $j - M_*$ scaling relations (Pulsoni et al. 2023) established for ‘ordinary’ ETGs. Pulsoni et al. (2023) measured the specific angular momentum of ETGs out to large radii using planetary nebulae.

For our dE sample we are in the position to correct the velocity curves for the inclination using the information derived from dynamical models (see Sec. 4.1). Fig. 6 shows both, inclination-corrected (filled circles) and uncorrected (open circles) estimates for the angular momentum of our dE sample. For LTGs with an exponential disk profile, a different approximation is used $j \approx 2 \cdot v_{max} \cdot r_{disk}$ with the disk scale length (e.g. Fall 1983; Scorza & Bender 1995; Romanowsky & Fall 2012). However, for LTGs j is also often inferred more directly from outer gas disks which extend to several r_e and enable inclination correction.

Fig. 6 shows, as has long been known, that *bright* ETGs and LTGs follow two near *parallel*, but *offset*, sequences (Fall 1983; Romanowsky & Fall 2012; Pulsoni et al. 2023). This so-called ‘Fall-relation’ is usually expressed in terms of stellar mass M_* and not M_B (see Fig. 27). It can be physically motivated from Λ CDM structure formation as its slope is remarkably close to what is expected from the angular momentum acquisition of dark matter halos, $j \propto M^{2/3}$ (Peebles 1969; Efstathiou & Jones 1979). Initially the gas (which later forms a galaxy’s stars) is expected to follow this halo relation, but the subsequent evolution, star formation, environment, etc. may modify how much is inherited from the stars (see also VW–II).

If we were to assume a magnitude-independent mass-to-light ratio, the theoretical Fall-relation $j \propto M_*^{2/3}$ translates to $\log_{10}(j) \propto -0.2667 \cdot M_B$ in magnitude space. In Fig. 6 we show linear fits to all LTG galaxies (blue) and to the ‘ordinary’ ETGs with magnitudes $\forall M_B \in [-18.5, -21]$ mag (black). Most of the latter were classified ‘Fast-rotators’ (see above Fig. 5) and they follow the halo-momentum relation very closely with $j \propto -0.2673 \cdot M_B$. The LTGs are not far off either with a slope of -0.2465 . The fact that LTGs are distributed on a slightly shallower curve may be because of the inclusion of dwarf LTGs in the fit. Dwarf spirals tend to fall above versions of the Fall relation fitted to massive spirals only (Chowdhury & Chengalur 2017; Butler et al. 2017; Kurapati et al. 2018). However, the full LTG sequence including dwarfs can still be described well by a single power-law with a slightly shallower slope (Posti et al. 2018; Mancera Piña et al. 2021).

In contrast, ETGs do not follow a single universal scaling relation. The most luminous (cored) ETGs tend to

⁸ For the ATLAS^{3D} ETGs we used their kinematic maps (Cappellari et al. 2011) and estimated v_m from the median of the bins around the maximum of the measured velocity.

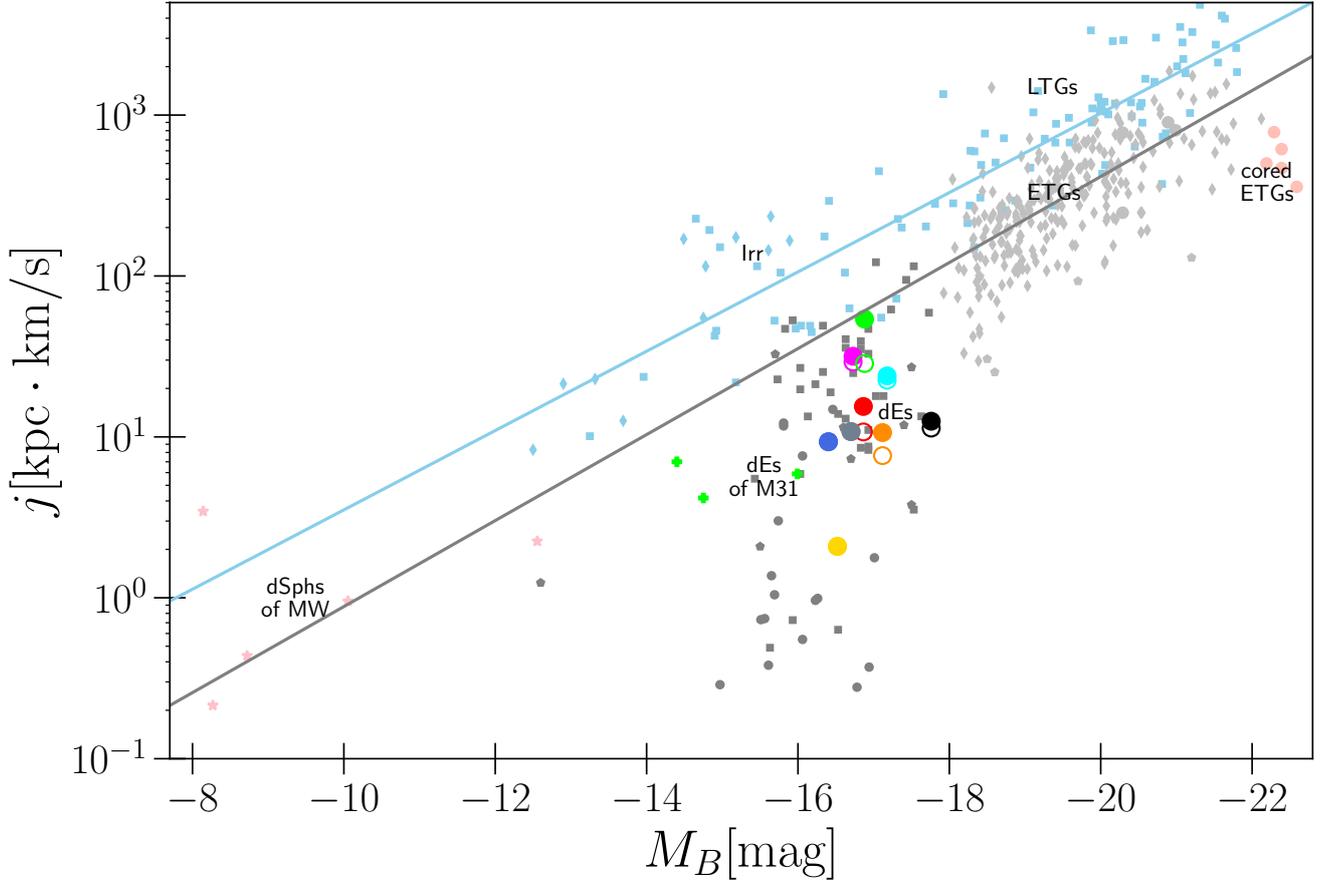


Figure 6. Estimate of the stellar specific angular momentum j versus absolute magnitude (B -band) for various galaxy types. For those studies where the magnitude was not stated we consulted Hyperleda (Makarov et al. 2014). For ETG samples we estimated j from the the stated velocity curves (see text), and when velocity data were noisy we clipped outliers. For the LTG samples we used the j directly from the source tables. *Blue diagonal:* Fit to all LTG points ($\log_{10} j = -0.2465 \cdot M_B + \text{const.}$). *Gray diagonal:* Fit to ‘ordinary’ ETGs with $M_B \in [18.5, 21]\text{mag}$ ($\log_{10} j = -0.2673 \cdot M_B + \text{const.}$). Massive ETG and LTGs ($M_* \gtrsim 10^{10} M_\odot$) are known to roughly follow the same relation ($j \propto M_*^{2/3}$), but LTGs are offset towards higher j (cf. Fall 1983; Bender & Nieto 1990; Romanowsky & Fall 2012). *Large, colored points:* Our Virgo dE sample (inclination-corrected: filled, not-corrected: open). *Small green crosses:* The dE satellites of M31 (NGC 147, NGC 185, NGC 205) inferred from individual stars (Geha et al. 2006, 2010) with velocity curves out to $\sim 8 r_e$. Studies of integrated light (smaller FoV) suggest significantly smaller velocities Bender et al. (1991); Simien & Prugniel (2002). *Gray squares:* Long-slit dE observations from Toloba et al. (2015). *Gray dots:* dEs from Geha et al. (2003) some of which have exceptionally low j (presumably because of a very small FoV or a slit not aligned with the kinematic major-axis). *Gray and light gray pentagons:* ETG sequence (dwarfs and ‘ordinary’ ETGs) of Bender & Nieto (1990). *Faint, red stars:* dSphs of the Milky Way from Martínez-García et al. (2021). For these, j is from *intrinsic* velocities as they derived their values from the Gaia proper motions and radial velocities of *individual* stars. *Light gray diamonds:* The ETGs of Emsellem et al. (2011). *Light gray dots:* power-law ETGs of Mehrgan et al. (2023). *Light, red dots:* Giant cored ETGs of Mehrgan et al. (2023). *Light blue diamonds:* Dwarf irregulars of Kurapati et al. (2018) (j is of stars only). *Light, blue squares:* Dwarf and massive spirals of Posti et al. (2018).

fall below the standard Fall-relation that fits the bulk of ‘ordinary’ ETGs much better. And, akin to what was noticed for the suppressed angular momentum parameter λ , nearly the entirety of dEs (and the dE satellites of Andromeda) fall *below* what one would expect from the Fall relation of ‘ordinary’ ETGs. At the low mass-end the two data points for the biggest Milky Way dSph (Fornax) still fall below the relation (similar to the dEs)

but the situation becomes more ambiguous for the smallest Milky Way dSphs which on average have a angular momentum in line with the Fall relation.

In parts this angular momentum suppression can be explained by the smaller mass-to-light ratios that dEs tend to have compared to ‘ordinary’ ETGs, but the dEs remain in tension with the standard Fall-relation as shown in the j vs M_* diagram (Fig. 27). The decrease

in j for the dEs could also be because the estimate for dwarfs may not be 1:1 comparable to that of the ‘ordinary’ ETGs, particularly if the total angular momentum of these galaxy types is distributed very differently in space. A large portion of the angular momentum of the dwarfs could ‘hide’ at larger radii. However, we argue the difference in j is too big to be explained by this *alone*. While the velocity profiles of our dEs only extend out to $\sim 1 r_e$ and likely still rise beyond that, none of our dEs show signs of a rapidly rising velocity profile. The velocity curves we found either start to plateau or rise mildly at the FoV edge (Fig. 3). In comparison the ‘ordinary’ ETGs, to which the Fall relation in Fig. 6 is fitted to, are actually much more suggestive of a rotation curve maximum that has not been reached yet and a detailed analysis of planetary nebulae tracers (Pulsoni et al. 2023) shows that these galaxies already hide a lot of angular momentum at large radii. This means for the dEs to fall on the same relation they would have to hide even more j at large radii than the ‘ordinary’ ETGs, all the while having lower angular momentum parameters λ_e in the centre.

To answer with more certainty whether dEs live below the standard Fall relation, rotation curves sampled out to much larger radii will be needed, e.g. by using non-stellar tracers as was done for the brighter ETGs. For the dEs of M31 the velocity profiles from stars alone were probed out to $8 r_e$ (Geha et al. 2006, 2010), which showed that rotation velocities reach their maximum significantly beyond $1 r_e$. However, even these higher velocities are still not sufficient to elevate the angular momentum j of these dEs to the level that is expected from the Fall relation of bright ellipticals (Fig. 6). For some dEs an analysis of their globular clusters suggests rapidly rising profiles beyond $1 r_e$. Beasley et al. (2009) for example have analyzed the two slow rotators VCC 1261 and VCC 1528 in our sample using globular clusters as kinematic tracers. Their results suggest much higher velocities which would be able bring the dEs onto the standard Fall relation. However, the reported uncertainties in the GC analysis are high and there is some tension with the stellar rotation profiles, which suggest a milder increase in velocity (Fig. 3).

In conclusion, while at least some of the initial angular momentum of dEs may have just been redistributed to larger radii (i.e. it is hidden at larger radii at $2 - 8 r_e$) we may still expect that a significant portion has been lost entirely. The processes responsible for this can be manifold and in superposition. They can be *internal* (e.g. supernova outflows) or *external* (e.g. galaxy harassment) as discussed below, but the comparison to the ‘ordinary’ ETGs suggests an onset (of importance) of

these mechanisms for galaxies with stellar masses below $\sim 10^{9.5} M_\odot$.

A comparison with studies of LTGs is also interesting. The finding that the average angular momentum parameter λ *increases* as one is moving along the sequence of ETGs with stellar masses from $10^9 M_\odot$ up to $10^{10} M_\odot$ appears to be traced even by the star-forming analogs (dwarf spirals). While the highest mass LTGs ($\gtrsim 10^{10} M_\odot$) have a relatively high λ that is more or less independent of the total galaxy mass (i.e. unlike ETGs with $\gtrsim 10^{10} M_\odot$), the amount of rotation seems to start to drop for LTGs with masses below $\log_{10}(M_*/M_\odot) = 9.3$ (Falcón-Barroso et al. 2019; Wang et al. 2020), similar to the low-mass behaviour in the ETGs sequence. In other words: the process responsible for the change in λ that occurs somewhere between $10^9 M_\odot$ and $10^{10} M_\odot$ could be a *common* one for both passive and star-forming galaxies. However, a comparison of the total angular momentum j in the LTG sequence shows this behaviour in λ is not accompanied by a similar behaviour in total specific angular momentum j . This may suggest that the suppression of λ for LTGs is more so because of a redistribution of the angular momentum to larger radii without changing J globally. This discrepancy between the behaviour of j in late-type dwarfs and early-type dwarfs could suggest that there is an additional mechanism removing the angular momentum entirely. We take up this discussion again in VW–II, where we analyse the relation between LTG and ETG dwarfs in conjunction with our results for their dark matter structure.

3.2. Correlations of angular momentum and stellar population with environment

Beyond the above dependence of the angular momentum on stellar mass, the environment could affect the amount of ordered motion in galaxies. In fact, it may even be the driving factor behind the suppressed j (and λ) of dEs (compared to ‘ordinary’ ETGs) we found, since they have lower stellar (or total) mass, their potential well and their ability to withstand their environment is reduced (Romero-Gómez et al. 2024). Similarly, the environment is often thought to be the main reason why dEs have stopped forming stars as they lost their gas reservoir due to external influences. In the following we investigate whether the angular momentum parameter and stellar population properties (age and metallicity) are a function of the galaxy environment.

For the purpose of a quantitative analysis of past environmental influences one must find a measure that quantifiably traces the degree to which each galaxy has experienced interaction with its host cluster. Commonly

used tracers are: The projected cluster-centric distance, the local galaxy number/luminosity density, distance to nearest large neighbour, or the local density of the intracluster medium. A priori it is not clear which tracer is best used as a proxy for past environmental influences. We argue one should treat all environment indicators with some caution. Firstly because of the large uncertainties in the distance along the line of sight and consequently its actual 3D position within the cluster. And secondly because a galaxy’s *current* location may not be very representative of the *past* interactions it had with the cluster and other galaxies. Some dEs might have experienced multiple pericenter passages already, or pre-processing in groups (Fujita 2004) which may not be reflected in its current day position. For the Virgo cluster this consideration might be especially crucial because it may be fairly young dynamically and still unrelaxed (Binggeli et al. 1987). Sybilska et al. (2017) have tested different environment proxies for a sample of Virgo dEs (including some of our sample galaxies). While the different proxies they tested may differ quantitatively, they agree at least qualitatively for the most part, i.e. if a strong correlation was present, it is detected in all proxies. Therefore we decided to use the most common and simple proxy for our study: The projected distance $\Delta r_{\text{sky},\text{M87}}$ to the central cluster galaxy of Virgo, M87.

Using this environment proxy we find a strong correlation of the angular momentum parameter λ with the Virgo environment (Fig. 7). The galaxies near the cluster center are slow rotators whereas galaxies in the cluster outskirts tend to possess more angular momentum. Similar results were found by Toloba et al. (2015) and Scott et al. (2020) who studied Virgo and Fornax dEs, respectively. Conforming with this environment correlation are also the dEs of the recently accreted galaxy group analyzed by Bidaran et al. (2020). As shown in Fig. 7 their angular momentum parameters tend to be higher than that of dEs deep in the cluster, although their $\lambda_{e/2}$ display a wide range of values which they attribute to the pre-processing of some of their dEs within the group.

This dependence of the rotational support vs environment may well extend from cluster scales down to the scales of individual galaxies. For example, within the Local Group, the dE satellites of M31 (which are a few magnitudes fainter than the Virgo dEs) that are closest to M31 show signs of tidal heating and reduced rotation due to their interaction with M31 (Bender et al. 1991; Geha et al. 2006, 2010). For the even fainter dSphs of the Milky Way, the correlation of angular momentum with environment also exists but is comparatively weak

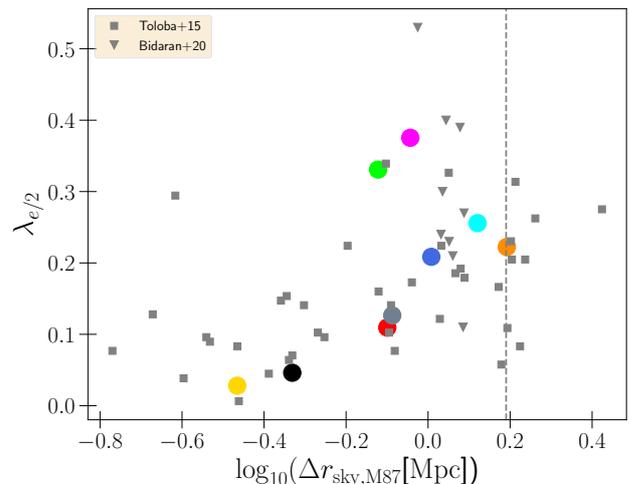


Figure 7. The angular momentum parameter λ within half the effective radius vs the environment proxy (the projected cluster-centric distance to M87). To convert the projected distance to Mpc we assumed a constant distance to Virgo of 16.5 Mpc, i.e. the position of the galaxy along the line of sight is not considered. The angular momentum parameter of dEs appears to decrease for the galaxies that have experienced stronger and longer-lasting interactions with the cluster environment, which confirms previous trends of angular momentum-environment relations in the Virgo-cluster (Toloba et al. 2015) and Fornax-cluster (Scott et al. 2020). Shown here are only the dEs of Toloba et al. (2015) and Bidaran et al. (2020) since these also inhabit the Virgo-cluster. *Dashed line:* The virial radius of the Virgo-cluster (Ferrarese et al. 2012)

(Martínez-García et al. 2021). We may expect that the fainter and less massive satellites of individual galaxies have very different dynamical time scales than the more massive dEs in galaxy clusters that infall into the cluster at a higher speed. Some of the cluster dEs may experience their first in-fall such that they only recently started to become heated by their environment, while for the small local satellite galaxies many close interactions with their host may have ‘washed out’ any correlation by now. Therefore the environment correlations we find in our sample may not be simply transferable to any environment or mass regime.

Contrary to the kinematics, there seems to be little to no correlation of the stellar population properties with the environment (Fig. 8). However, we do note that all dEs within the central Mpc (two-thirds of Virgo’s virial radius) are older⁹ than 6 Gyr while all dEs outside this radius are considerably younger with ages of 3 to 4

⁹ We measure the average age of both apertures. For VCC 1910 we adopt the ages from the literature (cf. App. B).

Gyr. The two youngest galaxies (VCC 308, VCC 2048) are the galaxies furthest away from the cluster center¹⁰ and both are located in a region with significantly lower projected number/luminosity densities than for the rest of the dEs (see Sybilska et al. 2017).

The question whether the population properties of dEs correlate with the Virgo environment is also debated in some of the existing population studies (e.g. Geha et al. 2003; Michielsen et al. 2008; Sybilska et al. 2017) that we discuss in App. B. Some find no correlation of environment with stellar populations while others find moderate correlations. In our dE sample it appears to be the case that within a projected cluster-centric distance of ~ 1 Mpc, the dEs can have a large variety in ages and stellar populations, but at larger radii (near the virial radius of Virgo around ~ 1.5 Mpc) the dE population is dominated by much younger ages (which was also noted by Michielsen et al. 2008). This could indicate that any correlation with environment becomes quickly saturated as the dEs experience their first infall into the cluster and their stellar populations become *homogeneous*. The young age of the galaxies outside the virial radius could imply that they have formed stars up until recently (see also Sec. 5.3), while the broader distribution of ages deep within the cluster center is simply a stochastic imprint of the epoch a galaxy was being quenched at. Instead of a continuous correlation with cluster distance, one would then expect a distinct cut-off at a certain cluster radius. If this radius is indeed in the vicinity of Virgo’s virial radius, our results favor a scenario where the driving quenching mechanism is a fast-acting mechanism like ram-pressure stripping. Such a process could already become efficient at the cluster’s virial radius and even at the first infall a galaxy experiences. Stellar population properties found in *Fornax* dEs seem to support this hypothesis more strongly than Virgo dEs, as Romero-Gómez et al. (2023b) found Fornax dEs exhibit no environment correlation with age and metallicity but a significant correlation with $[\alpha/Fe]$ that suggests at larger cluster-centric distances the star formation periods are more prolonged. A significant fraction of dEs may have also been quenched before arriving in the cluster due to pre-processing and, as they entered by the cluster, ram-pressure stripping triggered a short period of new star

formation (Bidaran et al. 2022; Romero-Gómez et al. 2024).

Environmental effects on a galaxy’s stellar population will be different at different radii. Spatially resolving potential gradients is therefore important. We quantify such a gradient as the log-linear change between the two apertures of the population properties that were shown in Fig. 8. In Fig. 9 we show these gradients for our dEs as a function of their environment and compare them to gradients from other IFU studies of Virgo dEs (Sybilska et al. 2017; Bidaran et al. 2023). Overall the age and $[Mg/Fe]$ gradients are fairly flat for our dEs and exhibit no discernible preference to positive or negative values, but almost all dEs have moderately more metal-rich centers. Only VCC 308, which is the dE most distant from M 87 and the only one classified as having an extended blue center (Tab. 1), and a few (but not all) of the *in-falling* dEs of Bidaran et al. (2023) show more noticeable gradients. The centers of these few galaxies are younger, more metal-rich and less α abundant. The location of these galaxies near Virgo’s virial radius suggests that as they fall into the cluster, the ram pressure exerted by the intra-cluster medium may have rejuvenated star formation in their centre (Boselli et al. 2022; Bidaran et al. 2023). For the galaxies that (presumably) entered the cluster a longer time ago there is no detectable correlation of the radial gradients with environment consistent with the analysis of Sybilska et al. (2017); Bidaran et al. (2023). The relatively flat age and abundance gradients suggests the galaxy’s main body is consistent with a nearly spatially homogeneous stellar population¹¹.

All together, the evidence for population-environment correlations in the literature are (especially in Virgo) still fairly tentative and exhibit, if any, only a moderate correlation. The inclusion of genuine field dEs in the population studies could help reveal a potential dichotomy of population properties as the global environment changes.

The more continuous and stronger correlation of the angular momentum with environment, on the other hand, suggests that the kinematic structure evolves much more gradually than the stellar population characteristics do. These results are consistent with dEs being transformed late-type dwarfs that were *quickly* quenched by their environment at the time of their first infall and, afterwards, were (more slowly) dynamically heated, losing their angular momentum in the process. In this formation scenario there exist a manifold of mecha-

¹⁰ Note that VCC 1261 has a large line of sight distance measurement (cf. Tab. 1). If correct VCC 1261 would be a further away from cluster center than the projected distance to M87 suggests. We have also checked whether the SSP properties are correlated with the 3D distance to M87 using the distance estimates in Tab. 1, but the population properties remained uncorrelated with environment.

¹¹ This does not hold for the small blue nuclei in the centre (cf. Paudel et al. 2011) which are not resolved in our study.

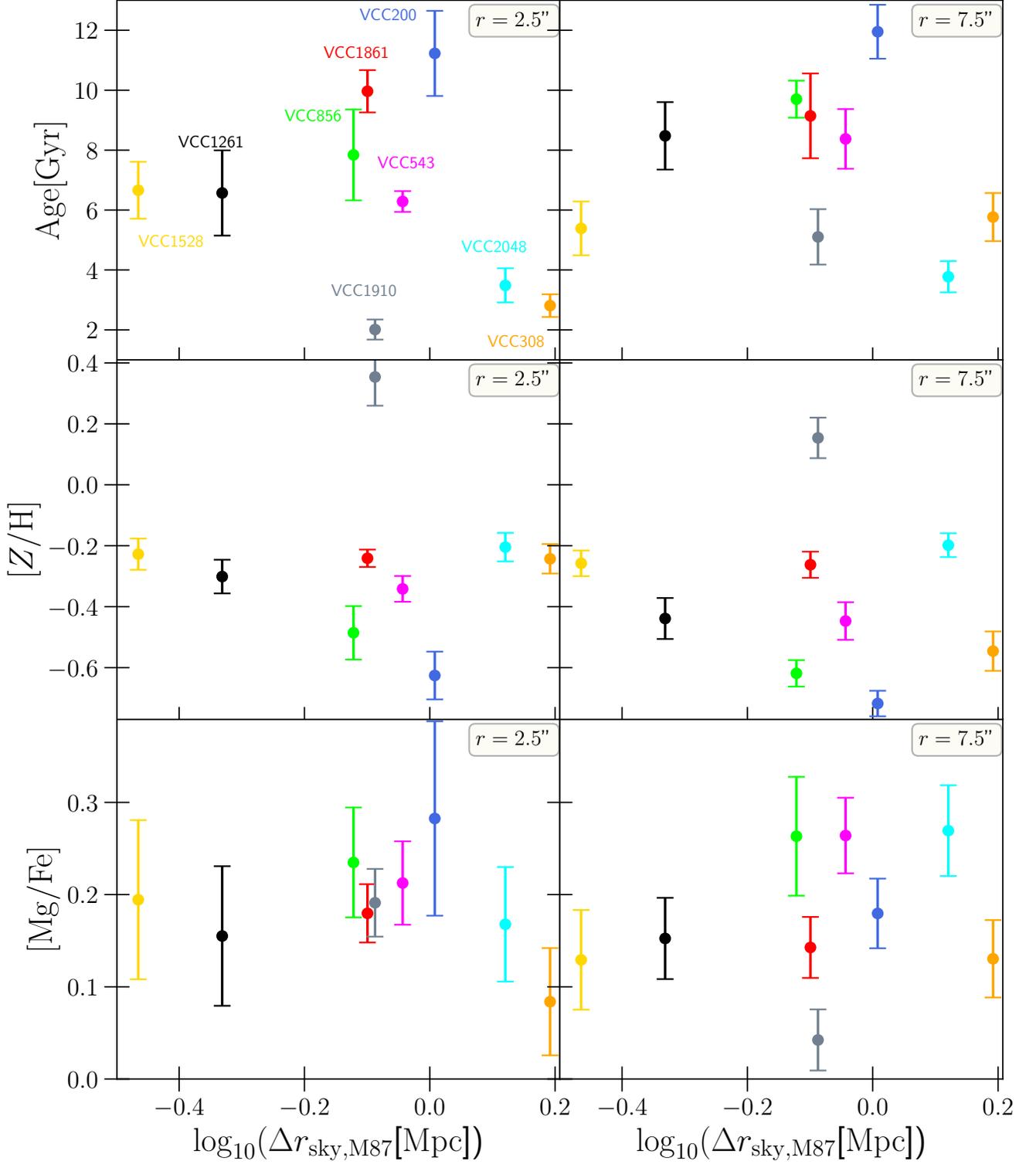


Figure 8. SSP age (*Top*), metallicity (*Middle*), and $[Mg/Fe]$ abundance ratio (*bottom*) extracted from the spectra at $r = 2.5''$ (*Left*) and $r = 7.5''$ (*Right*) versus the projected distance to M87 $\Delta r_{\text{sky},\text{M87}}$.

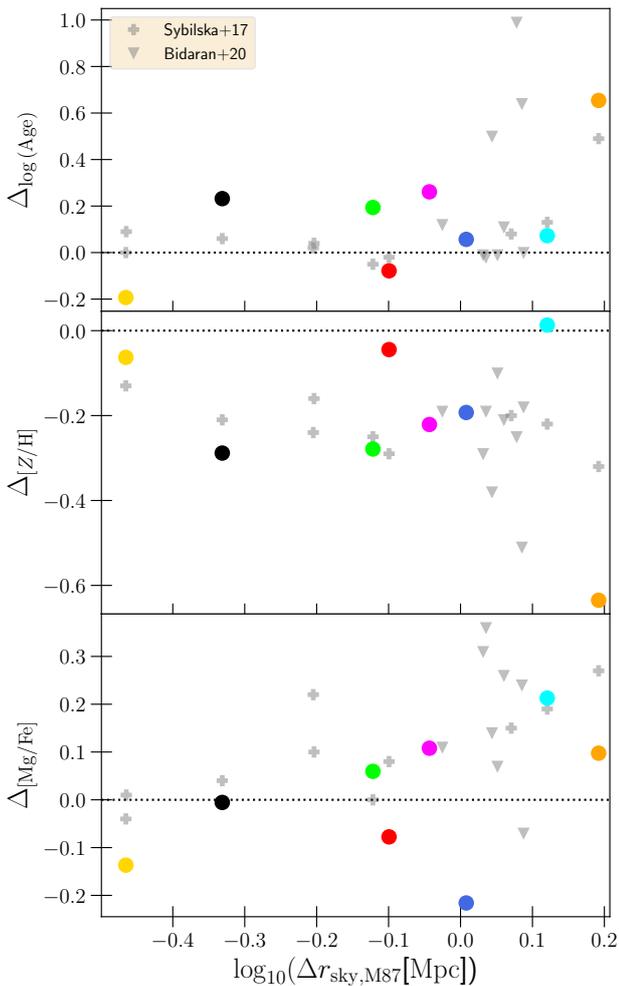


Figure 9. The log-linear population gradients versus environment for our sample together with the Virgo dEs of Sybiliska et al. (2017) and Bidaran et al. (2023). Both studies are based on IFU data and gradients are defined as the fitted slope of $\log(\text{age})$, Z/H , and $[Mg/Fe]$ within one stellar effective radius $\log(r_e)$. The gradients for our dE sample are measured as the change between two apertures we have spectra for, i.e. $\log(r = 7.5'')$ and $\log(r = 2.5'')$. VCC 1910 is excluded because of its poor inner spectrum.

nisms that can explain this behaviour when thought of in superposition. At larger radii the quenching mechanism keeps the stellar rotation relatively intact (e.g. starvation, ram-pressure stripping) but stops the star-formation quickly. But at later stages, when the dEs are deeper in the cluster, processes like tidal disruption and harassment by more massive galaxies heat up the stellar orbits, transforming the dE further. Furthermore, if the dEs are still able to form stars during their first infall, their remaining gas disks may become misaligned with the already existing stellar distribution which could also

lead to a dispersion dominated system as time progresses (Zeng et al. 2024).

However, we suspect that environmental effects are only a part of the broader picture. This is because environmental effects alone struggle to explain the dramatic increase of λ with total mass that occurs in the ETG sequence between $10^9 M_\odot$ and $10^{10} M_\odot$ (cf. Sec. 3.1) because of two issues: Firstly, the LTG sequence exhibits a similar sudden λ suppression in this mass regime (though not in j). Secondly, genuine field dEs that reside far outside any cluster also seem to display only low to intermediate angular momentum parameters (Janz et al. 2017) and are still quiescent. Both observations suggests that, regardless of external influences, low mass galaxies in and of themselves have a low angular momentum parameter and it is possible for late-type dwarfs to be quenched even in low-density environments. Perhaps the reduced potential well makes the dwarf galaxies more susceptible to *internally* induced dynamical heating and quenching, e.g. by supernovae winds or stellar bars, which causes an additional reduction of the angular momentum on top of the environmental influences.

4. CONSTRAINTS FROM DYNAMICAL MODELING

In the following we present the stellar mass and kinematic structure of the dE sample with the modeling setup we described in Sec. 2.4. We plot the line of sight kinematic profiles of the best AIC_p model we found for each galaxy together with the observed data in Fig. 3. The best orbit models fit the observations well and reproduce all discernible features in the first few Gauss–Hermite moments despite the fact that we fit the entire FoV as a whole (instead of quadrants individually). This suggests that our sample galaxies do not show obvious evidence for being non-axisymmetric. In fact, it is remarkable that the kinematic moments of all dEs, fast rotators *and* slow rotators, can be emulated this well by oblate axisymmetric models. In contrast, the most massive, slowly rotating ETGs often display photometric and kinematic signatures (e.g. Schechter & Gunn 1978; Williams & Schwarzschild 1979; Binggeli et al. 1985; Ene et al. 2018; Neureiter et al. 2023b) that require *triaxial* models and orbits to be emulated well. The fact that this is not necessary for any of the dEs may be another hint that they, unlike more massive ETGs, are indeed the remnants of transformed (oblate, axisymmetric) LTGs.

Fig. 10 presents the AIC_p-constraints for the outer and inner mass-to-light ratios and the viewing angles we obtained from all the orbit models that were calculated. Together this 3 parameter set fully describes

the 3D stellar mass distribution of the models $\rho_* = \Upsilon_*(\Upsilon_i, \Upsilon_o) \cdot \nu(i)$.

Instead of focusing only on the values of these nuisance parameters that are merely used to construct the 3D mass distribution, it is more meaningful to investigate the mass distribution they describe directly (see also argumentation in LT). Fig. 11 illustrates the (spherically averaged) densities of both the dark and stellar mass distributions of the *best* models we found for each galaxy, i.e. with $\Delta\text{AIC}_p = 0$ as shown in Fig. 10. We also indicate the relative importance of both mass components by displaying the cumulative dark matter fraction $f_{\text{DM}} = \frac{M_{\text{DM}}}{M_{\text{DM}} + M_*}$ against the radius in Fig. 11.

Apart from VCC 1261, the baryonic stellar mass of the dEs seems to be distributed quite differently from their dark matter. The stars dominate the center, but unlike the halo their density falls off much more steeply further outside, which results in the drastic and monotonic increase in the dark matter fraction. At least within the radial regime we analyse ($\lesssim 1 r_e$), the dark matter seems to play a secondary role. We will analyse and contrast the radial density gradients and flattening of both stellar and dark matter mass in a more detailed manner in the companion paper (VW–II). For the remainder of this work we focus on the interpretation of the recovered baryonic properties.

4.1. Viewing angles

We sampled 5 different inclination angles (and their corresponding deprojections) for each galaxy with the strategy described in Sec. 2.1. The recovered inclination angles (Fig. 10) appear to be strongly constrained. VCC 856 is the only galaxy where the viewing angle reaches the lower sampling limit at $i = 32^\circ$. Axisymmetric deprojections of this galaxy with a significantly smaller inclination are inconsistent with the observed ellipticity profile. Therefore it seems the orbit model of VCC 856 prefers to be as close to face-on as possible, which would be in line with the faint near face-on, spiral arm signature (given the arms are aligned with the equatorial plane of the galaxy’s main body spheroid). If the galaxy was closer to edge-on, we would not be able to see the pattern in the photometry. Likewise, the only other galaxy known to exhibit a hint of a visible spiral pattern, VCC 308, is also constrained to be fairly close to face-on (cf. App. E). This consistency of the dynamically recovered viewing angles with independent photometric signatures (the spiral arms are not explicitly accounted for in the orbit models) are strengthening our confidence in the accuracy and importance of the inclination recovery which we first tested on simulations in Lipka & Thomas (2021). The average *intrinsic* flat-

tening of the stellar mass distribution which is associated with the recovered inclination angle is shown later in Fig. 14 together with the intrinsic stellar kinematic structure. We will also compare the intrinsic stellar axis ratios q_* to the corresponding axis ratios of the dark matter component in VW–II.

4.2. Stellar mass-to-light ratio gradients

Dynamically derived stellar mass-to-light ratio gradients are scarce in the literature (e.g. Oldham & Auger 2018; Mehrgan et al. 2024), as mass-to-light ratios are often assumed to be spatially constant. However, even for the seemingly smooth ETGs, some studies suggest one may expect a spatial variation of Υ_* . For example, for giant ETGs, several stellar population studies indicate a variation of the IMF inside the effective radius (Ferreras et al. 2012; van Dokkum et al. 2017; Parikh et al. 2018, 2024) which entails a corresponding variation of Υ_* there. Using the same axisymmetric modeling technique as we use, Mehrgan et al. (2024) have detected a similar increase of Υ_* in the stellar bulges of 6 giant ETGs. All of them show a central peak in Υ_* with a gradient happening on a sub-kpc scale (more concentrated than previously anticipated).

These findings in the giant ETGs contrast with the dynamically derived *stellar* mass-to-light ratio gradients of the dE sample in our study (see also Fig. 15). The dEs exhibit considerable diversity, with some galaxies radially increasing in Υ_* , while others decrease. When calibrated in the z -band¹², the gradients of our dE sample are distributed around an average of zero.

Another difference with the large ETGs is that in the dE sample *all* stellar gradients are more moderate: the sample average of the *absolute* gradient is $\left| \frac{\partial \Upsilon_{*,z}[M_\odot/L_\odot]}{\partial \log_{10}(r[\text{kpc}])} \right| \approx 0.75$. The most extreme gradient we find is that of VCC 2048 with $\frac{\partial \Upsilon_{*,z}[M_\odot/L_\odot]}{\partial \log_{10}(r[\text{kpc}])} \approx +1.62$ (cf. Fig. 12). In contrast, the bulges of the giant ETGs in Mehrgan et al. (2024) have a much larger sample average¹³ of $\frac{\partial \Upsilon_{*,z}[M_\odot/L_\odot]}{\partial \log_{10}(r[\text{kpc}])} \approx -4.37$. Note though this average value is largely driven by two of the giant ETGs with extreme gradients of around -8 , while the rest of the sample is more moderate with gradients of about -2.5

¹² To calibrate the mass-to-light ratios to a common band we used a conversion factor derived from the stellar population models of Maraston (1998, 2005) assuming a Kroupa IMF and a range of metallicities and ages that are plausible for the dEs (Sec. 2.3).

¹³ The 3D gradients they found are stated in different bands but we approximately calibrated their values into the z -band for a comparison.

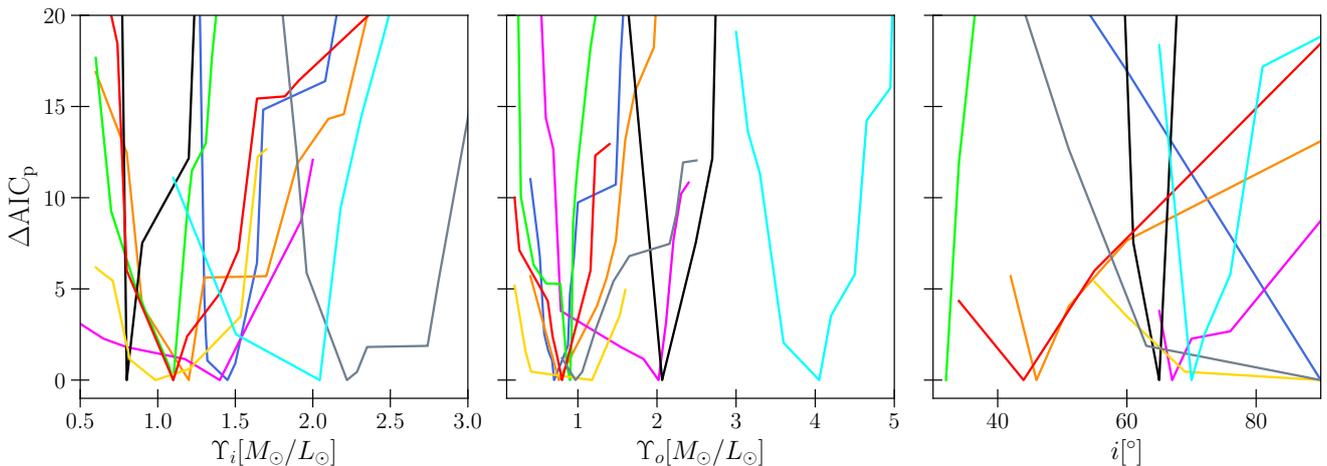


Figure 10. The ΔAIC_p envelopes for the intrinsic inner and outer mass-to-light ratios envelopes and the viewing angle of all the orbit models we probed for each dE. The stellar mass-to-light ratios are stated in their respective bands (Tab. 1). To obtain a realistic estimate of the errors we compared the differences in the best 25 models. This is roughly equivalent to $\Delta\text{AIC} \lesssim 10$ criterion.

It appears that dEs are distinct from the more massive ETGs in having much lower spatial variation with no systematic preference of rising or falling gradients. Since the stellar population properties (age and metallicity) are also approximately radially constant (Sec. 2.3), the stellar component appears to be well represented by a single homogeneous stellar population, at least at the scales we investigated (i.e. between 0.1 kpc to 1 kpc). However, we do note that the study of Mehrgan et al. (2024) was probing IMF variation in the very center of large ETGs. They choose to sample Υ_i and Υ_o at radii more concentrated in the center and not over the entire FoV like we did here. In other words, the gradients for both giant ETGs and dEs are measured between radii that are similar in physical scale (1 kpc scale) but relative to the overall galaxy size our gradients are sampled on a much more extended radial range. Perhaps one needs to probe the dwarf galaxies at much smaller scales to observe mass-to-light ratio gradients as strong as those found in giant ETGs. However, for the much smaller dEs this is currently unfeasible because our kinematic data does not resolve the very central parts well.

The inner and outer mass-to-light ratios Υ_i and Υ_o we sample with the orbit models determine the radial behaviour of the stellar mass-to-light ratio $\Upsilon_*(r)$ in 3D-space. In contrast, gradients recovered from a stellar population analysis (Sec. 2.3) describe the stellar population properties as observed on the sky. As such, mass-to-light ratios derived from populations are actually the ratios of the projected populations on the sky. The *projected* stellar mass-to-light ratios vs radius from both dynamics *and* population models are shown later in Sec. 5.

In projection the gradients appear even more moderate than they are in 3D space.

Despite these overall only moderate mass-to-light ratio gradients detected in the dEs, we still notice interesting relationships between the dynamical Υ_* variation and other properties. For instance, the two galaxies which exhibit distinctly strong and *positive* gradients, VCC 1261 and VCC 2048, also show distinctly different kinematic signatures in σ and h_4 , with them having strong central dispersion drops and negative h_4 -gradients (cf. Fig. 2). At the same time these two galaxies are also the galaxies with the largest measured velocity dispersion ($\sigma \sim 50 \text{ km s}^{-1}$). The dispersion drop could be a result of recent central star formation where the young and bright stars that were being formed in the center have not been heated up yet. This lowers both the central velocity dispersion and mass-to-light ratio compared to the older, dynamically hotter main part of the galaxy. Alternatively, it could be a signature of small objects, like star clusters or minor galaxies, that were accreted and disrupted within the galaxy’s extended envelope, increasing the random motions there.

This relation between the kinematic moments and the dynamically determined Υ_* -gradient seems to extend to the other galaxies: Fig. 12 shows the intrinsic 3D mass-to-light ratio gradients of the best dynamical models we found for each galaxy versus the observed velocity dispersion gradients. The latter were determined from the slopes of a simple linear regression fit to the radial dispersions (Fig. 3). As the Spearman correlation coefficient of $p_{\text{Spearman}} = 0.85$ indicates, the radial variation of the stellar mass-to-light ratio and the observed dispersion are strongly correlated. In fact if one were to

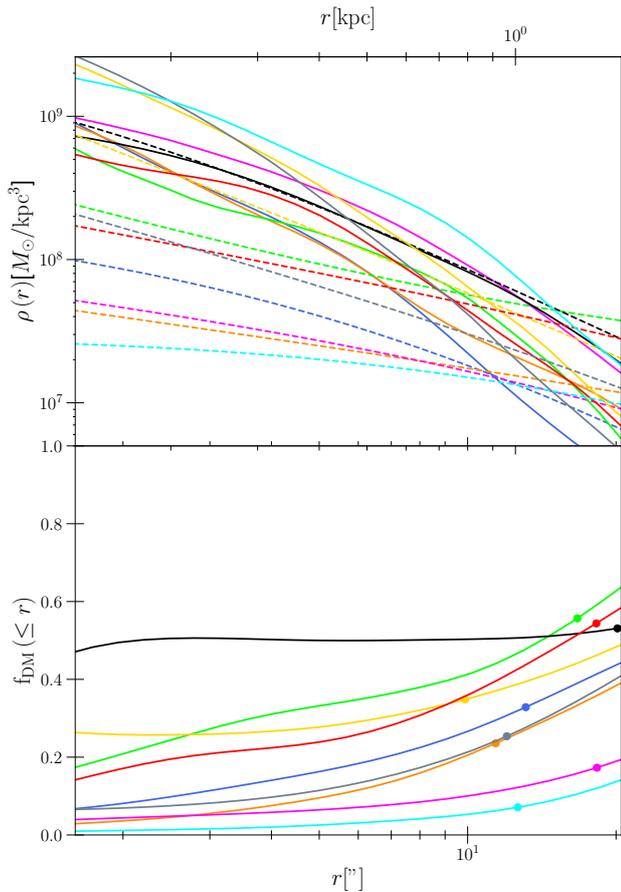


Figure 11. *Top panel:* The (spherically averaged) matter densities for the dE sample that we obtained from the best AIC_p model found for each dE. *Solid curves:* The stellar density (including the Υ_* gradients). *Dashed curves:* The corresponding dark matter halo density as modelled by the Zhao-parametrization (eq. 2). *Bottom panel:* The cumulative Dark matter fraction f_{DM} within radius r . The dots highlight the value of f_{DM} at one stellar effective, the typical radial extent of the kinematic data is $1.0 r_e$.

exclude VCC 856 (for which the classification as a dE and its kinematic recovery are questionable) from the analysis, the evidence for this correlation becomes even stronger with $p_{\text{Spearman}} = 0.95$. Interestingly the relation between the dispersion and Υ_* also seems to cross the zero point, i.e. galaxies with essentially flat dispersion profiles also have flat stellar mass-to-light ratios. It seems to be the case that the local stellar mass-to-light ratio and line-of-sight dispersion increase in lock-step. Such a positive correlation with the dispersion could naturally arise through the change in potential that comes with locally changing stellar mass-to-light ratio Υ_* . However, the dispersion gradients, particularly of VCC 1261 and VCC 2048, are very steep and confined to the central $5''$. The steep dispersion drops

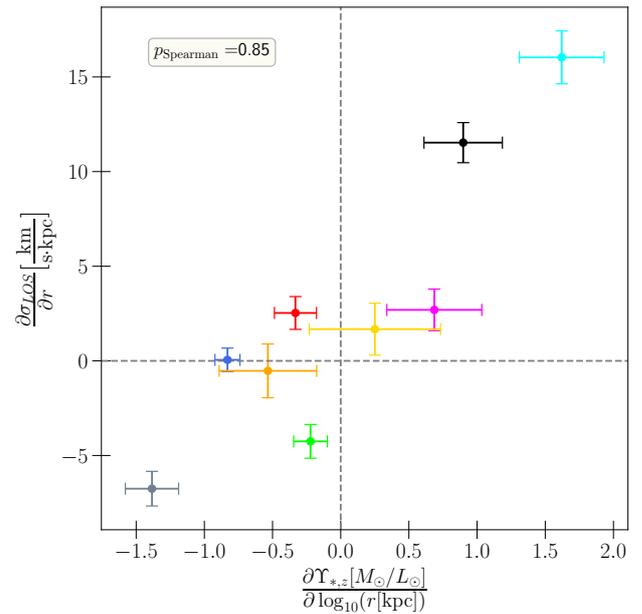


Figure 12. The observed average velocity dispersion gradients $\frac{\partial\sigma_{\text{LOS}}(r)}{\partial r}$, obtained from a linear fit to the observed radial dispersion profile, plotted against the 3D, log-linear, stellar mass-to-light ratio gradients of the dynamical models. The gradients were calibrated in the z -band for a consistent quantitative comparison.

of these two galaxies towards the centre (Fig. 3) could therefore be of different origin such as: the past accretion of smaller satellite galaxies, the presence and interaction with a bar, nucleus or embedded disk. However, none of these scenarios appears to be strongly supported by our analysis of the orbit structure that follows.

4.3. Intrinsic kinematic structure

The Schwarzschild models we employed yield constraints on the phase-space density of the stars. Therefore the 3D *intrinsic* kinematic structure of the stars and their associated velocity moments can be inferred from the kinematics of the individual orbits of the model and the recovered orbital weights (Thomas et al. 2004).

In spherical coordinates the anisotropy parameter β is a convenient quantity that describes by a single number the dynamical structure of the stars. It relates the second order velocity moments to one another as follows:

$$\beta = 1 - \frac{\sigma_\phi^2 + \sigma_\theta^2}{2\sigma_r^2} \quad (4)$$

where σ are the velocity dispersions in spherical coordinates (Binney & Tremaine 2008). A positive β means the structure is radially anisotropic, i.e. the stellar velocity dispersion is larger in the radial than in the tan-

gential direction. A negative β on the other hand implies the opposite.

Fig. 13 illustrates the spherically averaged anisotropy of the best dE models as a function of radius together with other ETG samples. Apart from VCC 856, the only genuine fast-rotator (Fig. 5) and an outlier in many regards (App. E), all dEs have a relatively homogeneous, isotropic orbit structure at most radii, with signs of mild radial anisotropy in the center ($< 0.3 r_e$). At large radii near or outside the FoV and within scales smaller than the central spatial resolution limit, the profiles diverge more. However, this is expected as the inferred orbital motions at these scales become less reliable due to the lack of kinematic data constraints in these regions (see also LT).

Despite the phenomenological diversity in the observed Gauss–Hermite moments (Fig. 3), the intrinsic orbit structure of the galaxies seems fairly homogeneous. This lack of spatial anisotropy variations implies a more direct connection between the observed dispersion profiles (Fig. 3) and the corresponding intrinsic *mass* profiles, consistent with the strong correlation between mass-to-light ratio gradients and velocity dispersion gradients (Fig. 12).

An orbit structure that is essentially isotropic throughout with a mild radial anisotropy in the center is remarkable. In contrast, the β profiles of ‘ordinary’ ETGs can be quite heterogeneous. Power-law ETGs ($\log(M_*/M_\odot) \sim 10 - 11$) may on average be nearly isotropic (Cappellari et al. 2007) but their profiles can vary strongly with radius and are generally far from isotropic at any point (e.g. Thomas et al. 2007, 2009b, 2014; Santucci et al. 2022). The even more massive cored ETGs are distinct again. They are found to be tangentially biased in their center and strongly radially biased further out (e.g. Thomas et al. 2014, 2016; Mehrgan et al. 2019), which is believed to be a result of a black hole core scouring process (cf. Kormendy & Bender 2009; Kormendy et al. 2009; Kormendy & Bender 2013; Thomas et al. 2014; Rantala et al. 2018). The scouring may even be a protracted process (Frigo et al. 2021) which could also result in intermediate states being observed where the core is not yet tangentially biased (e.g. Neureiter et al. 2023b). Conversely to both types of ‘ordinary’ ETGs (cored and power-law) the structure of the dEs is quite isotropic and homogeneous. Furthermore the tendency to be mildly radially anisotropic in the centre appears to be unique to the dEs.

This orbit structure seems to suggest that the evolution of dEs is distinct and again we note the ‘trichotomy’ across the stellar mass sequence from $10^9 - 10^{12} M_\odot$ (cf. Sec. 3.1). However, we do caution that a comparison be-

tween the two is not straightforward because we probe different scales relative to the galaxies’ sizes. The massive cored ETGs in Fig. 13 indicate that the stars typically are on more tangential orbits within about 0.5 kpc and become radial beyond 1.5 kpc. However, the physical scale involved here is the sphere-of-influence radius of the central supermassive black hole which happens to be around 0.1–1 kpc for most of the massive core galaxies studied (Thomas et al. 2016). In the dEs, however, we do not resolve the sphere of influence of any black hole, i.e. we are not able to detect any scouring-related effects in our sample galaxies.

The nearly isotropic structure (with mild radial anisotropy in the center) may be more akin to that of the less massive dwarf spheroidal galaxies (dSphs) and the dEs in the Local Group for which the anisotropy was studied. For example, De Rijcke et al. (2006) analyzed the velocity dispersion tensor of the 3 dE satellites of M31 based on slit kinematics and concluded all 3 galaxies are fairly isotropic within the central 0.8 kpc, the region that is most constrained by their data. Kowalczyk & Lokas (2022) used Schwarzschild models for the *resolved* kinematics of the Fornax dSph and found an anisotropy profile that remains fairly close to isotropy at most radii with signs of radial anisotropy in the very center. Jardel & Gebhardt (2012); Jardel et al. (2013) found β profiles for the dSphs Fornax and Draco that are slightly radial yet close to $\beta = 0$ with only a small increase in β throughout the relevant radial range. Using a Jeans modeling technique, De Leo et al. (2023) found a very radially constant and close to isotropic orbit structure for the Small Magellanic Cloud after removing tidally disrupted interlopers from their analysis. The finding that the Milky Way’s dSphs are fairly isotropic at all radii may be explained as a consequence of their tidal interactions with the Milky Way. N-body simulations of dSphs in a Milky Way like potential suggest that ‘tidal stirring’ (Mayer et al. 2001) could have transformed the orbit structure of initially disk dwarf galaxies to become fairly isotropic as a result of several interactions with the Milky Way (cf. Klimentowski et al. 2009; Lokas et al. 2010).

If this effect of isotropic transformation of dSphs via tidal interactions can be transferred to the larger scales of the more massive dEs, then the isotropic structure we find for the dEs is not necessarily their primordial state but a result of processing late-type progenitors via interactions with the cluster environment. It could also explain why VCC 856 does not conform to the homogeneous β structure of the rest of our sample, as it has not yet been fully transformed (or is in the process of) as evidenced by the face-on spiral arms found in VCC

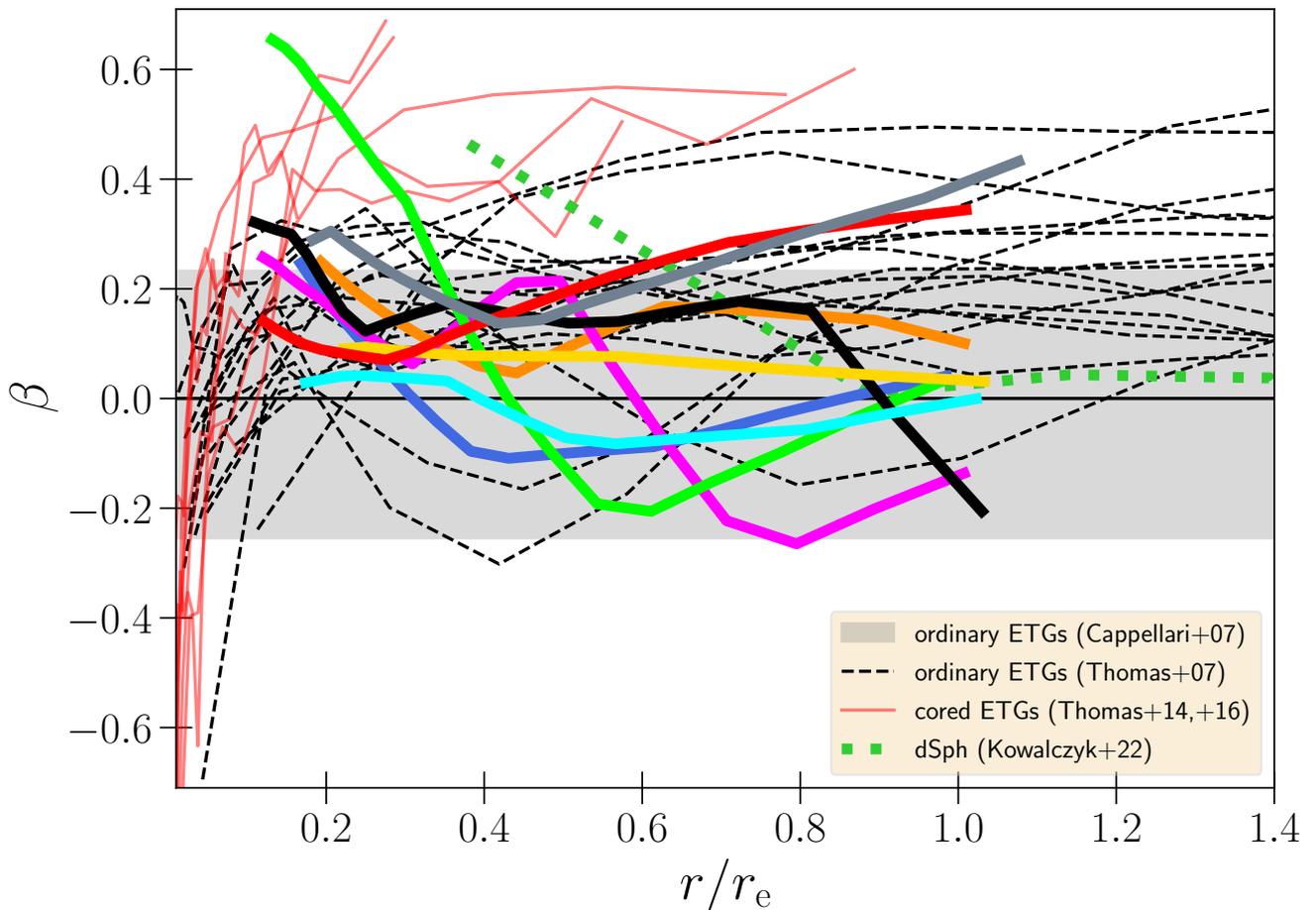


Figure 13. The spherical anisotropy parameter β (eq. 4) vs radius scaled by the effective radius. The typical size of the FoV for our dE sample is about $1.0 r_e$, and outside of the FoV the anisotropy is not well constrained. The same is the case for radii smaller than the spatial resolution limit. Using our error estimation criterion (Sec. 2.4) we find a typical 1σ error of the anisotropy parameter of $\Delta\beta = 0.065$ (within 1 kpc), with VCC 856 being the most uncertain at $\Delta\beta = 0.11$. *Red curves:* The anisotropy profiles of giant cored ETGs (Thomas et al. 2014, 2016). Due to black-hole scouring these have formed tangentially anisotropic cores ($\lesssim 0.05 r_e$) within a radially anisotropic envelope. *Black dashed lines:* ETGs from the sample of Thomas et al. (2007) which mostly consists of ‘ordinary’ ETGs with magnitudes within $M_B \in [-18.8, 22.6]$ mag. *Grey band:* The 90th percentile of the anisotropies of the 23 ‘ordinary’ ETGs of Cappellari et al. (2007). Note that these values are radial averages. If the orbital structure changes from tangential to radial (as is often the case), then this averaging can make the galaxy appear closer to $\beta = 0$ than it actually is. *Green, dotted line:* The Fornax dSph from Kowalczyk & Lokas (2022).

856. Tidal stirring would also be a natural explanation why dSphs and dEs tend to be predominantly pressure-supported and with low or moderate angular momentum (Fig. 5) specifically in the center of Virgo (Fig. 7). While the tidal stirring hypothesis may be successful in explaining the isotropic structure, the fact that our dEs inhabit fairly different environments within the Virgo cluster, and yet they all seem to have a similar isotropic structure, is peculiar. We would expect the dEs in the cluster center to have experienced on average more tidal passages, in which case we may expect a correlation of β with the cluster environment which we do not seem to find.

Alternatively, if the isotropic structure we find is not a result of interactions with the cluster, it may stem from an *internal* secular process that has driven the kinematic structure towards a specific configuration. In large ETGs the effects of these internal processes may be obscured by the more violent changes induced by mergers.

We found the *spherical* anisotropy parameter β to be almost constant with the radius and very homogeneous across all dEs in our sample. However, this impression of a completely isotropic stellar structure changes when looking at the anisotropy of the second velocity

moments¹⁴ in cylindrical coordinates defined as:

$$\beta_z = 1 - \frac{\langle v_z^2 \rangle}{\langle v_r^2 \rangle} \quad (5)$$

quantifying the anisotropy in the meridional plane. And in the corresponding sagittal plane:

$$\gamma = 1 - \frac{\langle v_\phi^2 \rangle}{\langle v_r^2 \rangle} \quad (6)$$

Fig. 14 shows these anisotropies (averaged within a sphere of radius 0.8 kpc $\sim 10''$) versus the average *intrinsic* ellipticity ϵ_{intr} . For our dE sample both anisotropy parameters appear to be correlated with the intrinsic flattening. While the near spherical dEs are essentially isotropic, the velocity ellipsoid deforms as the stellar distribution becomes more flattened. The positive β_z suggests that flattened dEs have a significantly smaller kinetic energy perpendicular to the equatorial plane, while the negative γ suggests a relatively large azimuthal energy (only some of it comes from ordered motion).

When compared to the angular momentum parameter of the galaxies (Fig. 5) we find that most galaxies with a higher angular momentum parameter $\lambda_{e/2} > 0.1$ also have a larger velocity dispersion in the equatorial plane ($\beta_z \sim 0.3$) suggesting both pressure and rotation support play a role in their flattening. There are exceptions, like VCC 1261, which is mildly flattened but has essentially no angular momentum (at least within $r_e/2$) both on the sky and also intrinsically in terms of ordered motion. Instead its flattening seems entirely supported by its relative lack of velocity dispersion in the z -direction.

In Fig. 14 we also show the cylindrical anisotropies of ‘ordinary’ ETGs (Cappellari et al. 2007; Thomas et al. 2009b) which were obtained with axisymmetric Schwarzschild models. While the flatter dEs overlap with the ‘ordinary’ ETGs, they trace the upper boundary in β_z and the lower boundary in γ of the ETG distribution. This may suggest that dEs are not heated like many of the most massive ETGs which have likely experienced mergers that would erase ordered motion more chaotically. Especially dry mergers are expected to make the orbit structure radially anisotropic, which is noticeable in a higher, positive γ (Thomas et al. 2009b). In comparison our dEs have low, negative γ which makes (dry) mergers unlikely. Instead the shape of their velocity ellipsoid is still very much aligned with the ori-

entation of a potential disk progenitor. This is compatible with the scenario that dEs stem from quenched LTG progenitors. A natural explanation for the large β_z is that their kinetic energy is a relic of dissipation by a gaseous disk before the dEs were being quenched. For the flattened dEs, imprints of the disk structure remained in the orbit structure until today, while for the large fraction of near-isotropic and spherical dEs, the heating process has come to its conclusion expunging any traces of its progenitor and its net angular momentum.

Orbit modelling of ETGs, in a similar mass range as our dEs, by Ding et al. (2023) has suggested that the environment (cluster infall time) affects the orbits a galaxy’s stars occupy. Galaxies that fell into a cluster recently have a higher fraction of orbits with high z -axis angular momentum than galaxies that entered the cluster long ago. However, contrary to the case of the angular momentum parameter $\lambda_{e/2}$ (cf. Fig. 7), we do not find any strong correlation of the anisotropy and intrinsic velocity moments with the 2D (or 3D) distance to M87. Instead we find examples of flattened/anisotropic dEs but also spherical/isotropic dEs in both the cluster center and in the periphery.

All in all, dispersion support appears to be a non-negligible contributor to the dE flattening. Compared to ‘ordinary’ ETGs the dEs have a lower amount of ordered motion (Sec. 3.1), yet their anisotropy structure (high β_z and low γ) that is not correlated with environment suggests their orbits are still very much aligned with the potential progenitor disk. It is not obvious why the ordered motion (i.e. angular momentum) of dEs should be diminished by the environment more effectively, while the higher kinetic energy in the equatorial plane is mostly preserved. One would expect tidal interactions that reduce the ordered motion to increase σ_z and, hence, to reduce β_z . Instead, it may be more plausible that the suppressed angular momentum of dEs is not a result of environmental processing, but simply a feature of their different gravitational assembly. We revisit this scenario in VW–II where the dark matter constraints allow further insights into the assembly conditions of dEs.

5. THE IMF OF DWARF ELLIPTICALS

Studies of the Milky Way and local late-type galaxies convey the picture of a *universal* (Kroupa or Chabrier) IMF (Kroupa 2001, 2002; Chabrier 2003; Brewer et al. 2012), yet the circumstances seem more complicated for early-type galaxies (e.g. van Dokkum & Conroy 2010). For the most massive ETGs the results often suggest a more Salpeter-like IMF (Salpeter 1955) and evidence

¹⁴ In the axisymmetric models the second velocity moments equal the velocity dispersions in all but the azimuthal direction as it includes ordered motion as well.

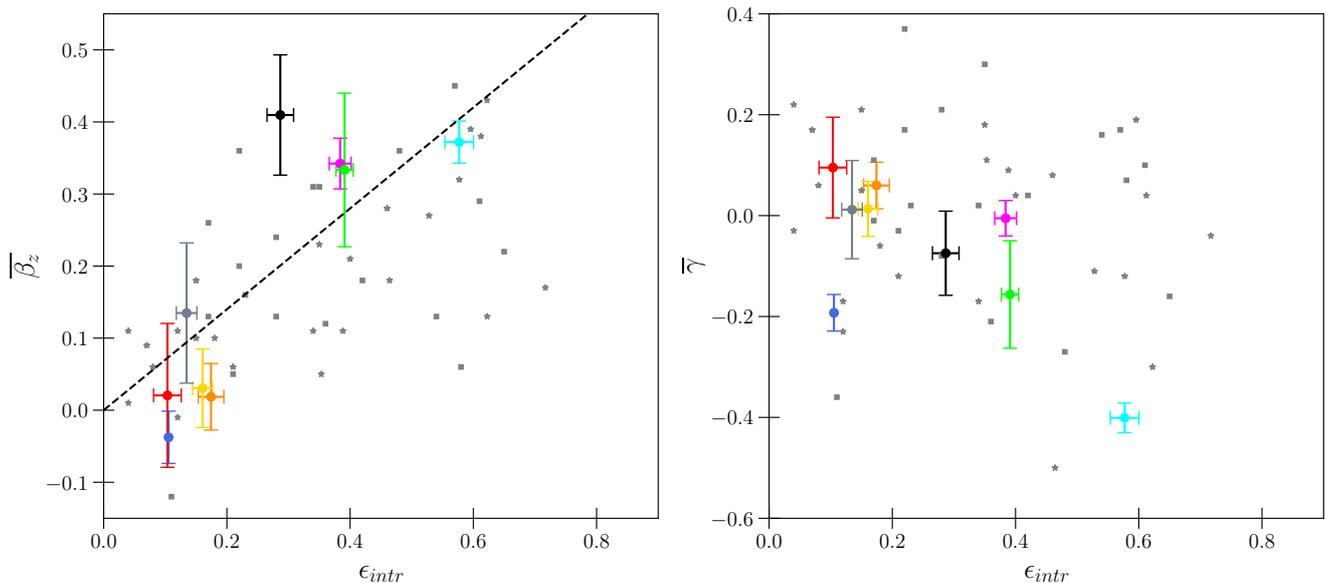


Figure 14. The average *cylindrical* anisotropies β_z (left) and γ (right) vs the *intrinsic*, average stellar ellipticity ϵ_{intr} . The anisotropies were averaged within the sphere of radius $r \leq 0.8$ kpc ($\sim 10''$) because values outside the FoV are poorly constrained by the data. *Squares*: The massive COMA ETGs of Thomas et al. (2009b). *Stars*: The ETGs from Cappellari et al. (2007). The *dashed* diagonal line indicates the empirical relation $\beta_z = 0.7 \cdot \epsilon_{intr}$ from Cappellari et al. (2007), where flatter galaxies have higher anisotropy β_z .

is mounting that the IMF varies with radius, age, or metallicity, and abundance ratios (e.g. Smith et al. 2012; La Barbera et al. 2013; van Dokkum et al. 2017; Parikh et al. 2018). Determining whether the IMF of the dEs behaves more like that of the large ETGs or that of LTGs may allow us to distinguish between the formation scenarios responsible for the dEs.

In our study we can compare two *independent* measurements of the stellar mass-to-light ratios for each dE obtained from one and the same set of VIRUS-W spectra which allows us to probe the validity of an assumed IMF. The first mass-to-light ratio measurement we obtained is from the dynamical modeling (Sec. 4). The second set of mass-to-light ratios stems from the stellar population analysis (Sec. 2.3) where we assume a certain form for the IMF as a reference. We decided to use the Kroupa IMF as reference and calculated the present day stellar mass-to-light ratios Υ_{Kroupa} from the SSP analysis for each of the analyzed spectra (i.e. at $r = 2.5''$ and $r = 7.5''$). In the following, unless stated otherwise, we specify Υ_{Kroupa} in the same bands we used for the dynamical measurements (cf. Tab 1). One can then compare the Υ_{Kroupa} with the dynamically derived mass-to-light ratios and conclude whether the assumption of a Kroupa IMF is accurate or if the IMF should be more *sub-Kroupa* or *super-Kroupa*.

In Fig. 15 we juxtapose the dynamical Υ_{dyn} and the two corresponding population mass-to-light ratio

Υ_{Kroupa} which were obtained from the two spectra in the annuli centered around $r = 2.5''$ and $r = 7.5''$. For reference we also show the *total*, *3D* dynamical mass-to-light M_{tot}/L (i.e. including the dark matter, but excluding the SMBH) and the projected mass-to-light ratio obtained from a simple ad hoc dynamical modeling where mass follows light, i.e. a dynamical model with neither dark matter, nor a stellar gradient.

As in other galaxy samples, we find there can be significant discrepancies between the dynamical and population mass-to-light ratios for some of the sample galaxies (e.g. Thomas et al. 2011b; Cappellari et al. 2013a; Posacki et al. 2015; Mehrgan et al. 2024). Using the relative differences $\Delta_* = (\Upsilon_{dyn} - \Upsilon_{Kroupa})/\Upsilon_{mean}$ with $\Upsilon_{mean} = (\Upsilon_{dyn} + \Upsilon_{Kroupa})/2$ we find values up to $\Delta_* = \pm 1$. An often used alternative to the IMF parameter Δ_* as defined here is the mass normalization parameter $\alpha_{IMF} = \Upsilon_{dyn}/\Upsilon_{Kroupa}$. While the use of α_{IMF} does not change any of the following conclusions we give values of this alternative parameter in Tab. 2.

Fig. 16 shows the IMF parameter and the SSP age for our dE sample together with the ‘ordinary’ ETGs of the ATLAS^{3D} survey (Cappellari et al. 2011) scaled by the central velocity dispersion (which is a proxy for the total mass).

For giant ETGs the velocity dispersions are known to be positively correlated with the IMF parameter (e.g. Cappellari et al. 2013b; Cappellari 2016; Zhu et al.

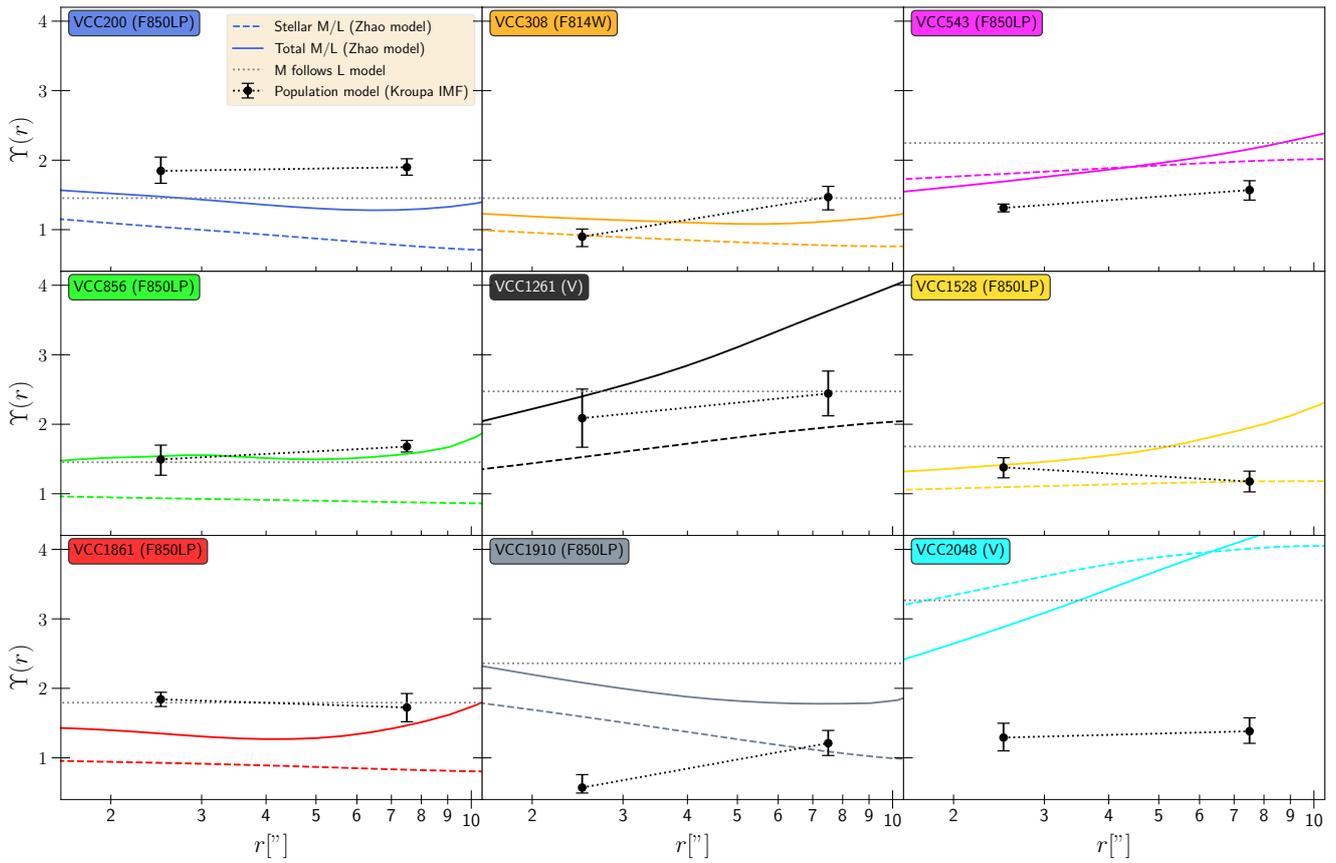


Figure 15. Comparison of the local mass-to-light ratio Υ as a function of radius r for all Virgo dEs as obtained from different models. Each panel shows one dE, with the box showing its VCC-id and the band in which the mass-to-light ratio is measured (cf. Tab. 1). The *colored* lines show the mass-to-light ratios we derived using dynamical modeling: *Dashed*: The projected Υ_* of the dynamically decomposed stellar component. *Solid*: The total 3D dynamical mass-to-light ratio Υ_{tot} (i.e. including the dark matter). Note that Υ_{tot} is not a projected quantity, as this would include all the dark matter along the LOS, i.e. even the poorly constrained DM far outside a galaxy’s FoV. This can make the *projected* Υ_* appear to be higher than the 3D Υ_{tot} values at some radii, even though the latter is by definition the upper bound for the 3D stellar mass-to-light ratios. The *gray, dotted horizontal* line shows the mass-to-light ratio of the best mass-follows-light model (i.e. a dynamical orbit model with no dark matter and stellar Υ -gradient). The *black dots* with errorbars show the stellar mass-to-light ratio from the population analysis (Sec. 2.3) in the two annuli centered around $r = 2.5''$ and $7.5''$. We also draw a connecting line between the two points to highlight a population gradient if it exists.

2024). At the high mass end ETGs are almost all very old and their average IMF is Salpeter-like¹⁵. But as the total galaxy mass decreases, galaxies often are younger as well and the typical IMF becomes progressively lighter, reaching Kroupa levels (i.e. $\Delta_* = 0$). However, analogous to the break in the angular momentum parameter λ at $\log(M_*/M_\odot) \sim 10.5$ (or equivalently $\log(\sigma) \sim 2.1$), there appears a break in these ETG correlations. Above this mass threshold, Δ_* and the age are positively correlated with the dispersion and show relatively little scatter. Below the threshold the scat-

ter between different galaxies increases dramatically and the IMF and age become independent of total mass and dispersion.

In this view our dEs are indistinguishable from those ‘ordinary’ ETGs with dispersions below $\log(\sigma) \sim 2.1$. While our dEs can be both, substantially super-Kroupa or sub-Kroupa, they are within the large scatter of the ‘ordinary’ low dispersion ETGs and they have Kroupa or Chabrier (i.e. just below Kroupa) IMF on average. A similar result was found by Tortora et al. (2016) who compared the IMF parameter of the dEs sample of Toloba et al. (2014) with ‘ordinary’ ETGs. Their dEs also exhibit remarkable diversity, suggesting both super- and sub-Chabrier IMFs, with the average dE being consistent with a Chabrier IMF. This diversity they mea-

¹⁵ Though this changes if spatial IMF gradients are considered (Mehrgan et al. 2024).

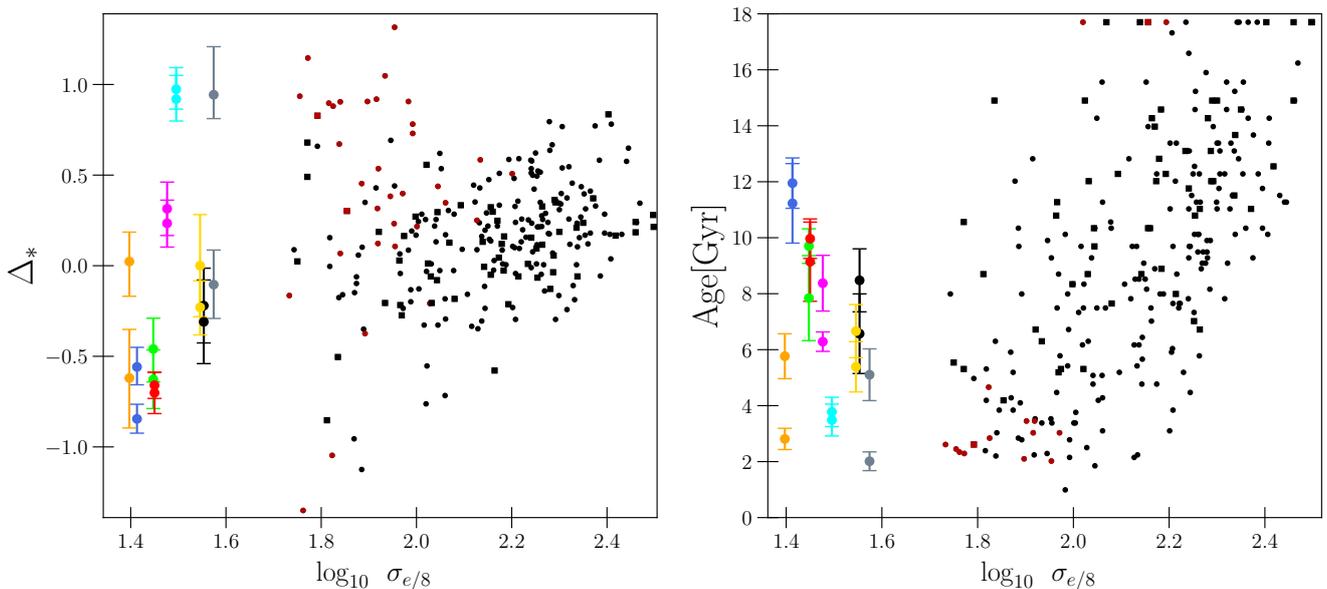


Figure 16. A comparison of the IMF parameter (*left*) and the single stellar population age (*right*) of dEs with those of the ‘ordinary’ ETGs from the ATLAS^{3D}-survey (Cappellari et al. 2011). All galaxies are shown versus their central velocity dispersion (measured within $r_e/8$) which is essentially a proxy for their total mass. ATLAS^{3D} ETGs that reside in Virgo are highlighted using squares, the others are shown as circles. We calculated the Δ_* for a Kroupa IMF by transforming the mass-to-light ratios from Cappellari et al. (2013b) which were given for a Salpeter IMF. The SSP equivalent age of the ATLAS^{3D} galaxies was obtained from McDermid et al. (2015) and is stated within a $1 r_e$ aperture. For our dEs we show the results for both apertures that we analyzed. Cappellari et al. (2013b) and McDermid et al. (2015) flag their galaxies depending on the quality of the data and analysis. For details on the definition of these flags see the corresponding studies. We show galaxies flagged with the best quality in *black* and ‘bad’ galaxies that may not be as reliable in *red*.

sured could have been due to the more simple, restrictive dynamical mass estimates they employed (spherical, isotropic Jeans equations), but the much more advanced Schwarzschild models employed in our work suggest a similar degree of IMF diversity.

This substantial scatter of galaxies below $\log(\sigma) \sim 2.1$ poses the question whether it stems from a real non-universality of the IMF or is simply due to an increased statistical uncertainty in the dynamical or population modeling. If the IMF scatter of the low mass galaxies is a result of a non-universal IMF we may hope to identify a physical reason for changes in the IMF by looking at correlations of Δ_* with other properties. The strongest correlations we find are with the SSP age and metallicity. It appears that the older and the more metal-poor a stellar population is, the larger its Υ_{Kroupa} is relative to Υ_{dyn} , which results in a smaller Δ_* . This correlation with the stellar population properties is illustrated in the *left* panel of Fig. 17 which displays the differences Δ_* versus the SSP ages we derived from the VIRUS-W spectra. For comparison we also show the points one would get for Δ_* if one would assume the ages and metallicities published in the literature (cf. Fig. 24). For the literature values we recalculated the mass-to-

light ratios Υ_{Kroupa} under the assumption of a Kroupa IMF.

A similar Figure for the correlation with metallicity is shown in App. D. Since the age of the dEs is anti-correlated with metallicity (Fig. 4) the $\Delta_* - [Z/H]$ relation follows as a corollary. We can expect the average metallicity background for each cycle of star formation to increase with the formation epoch. This naturally changes the initial conditions for star formation over time and, possibly, the form of the IMF (e.g. Li et al. 2023). In the following sections we explore whether this correlation of the IMF parameter with age (and/or metallicity) is real and, if so, what could explain it.

5.1. Are the dynamical models robust?

The mass-to-light ratio discrepancies Δ_* could originate from an erroneous *dynamical* stellar mass-to-light ratio estimate that is caused by a large statistical uncertainty in the dynamical decomposition of dark and stellar matter in which case the Υ_{dyn} would not be very representative of the actual mass bound in stars. One could even assume the worst case where the mass decomposition is completely random and the Υ_{dyn} for the different dEs randomly scatter around some average value depending on how much dark matter mass is mistakenly

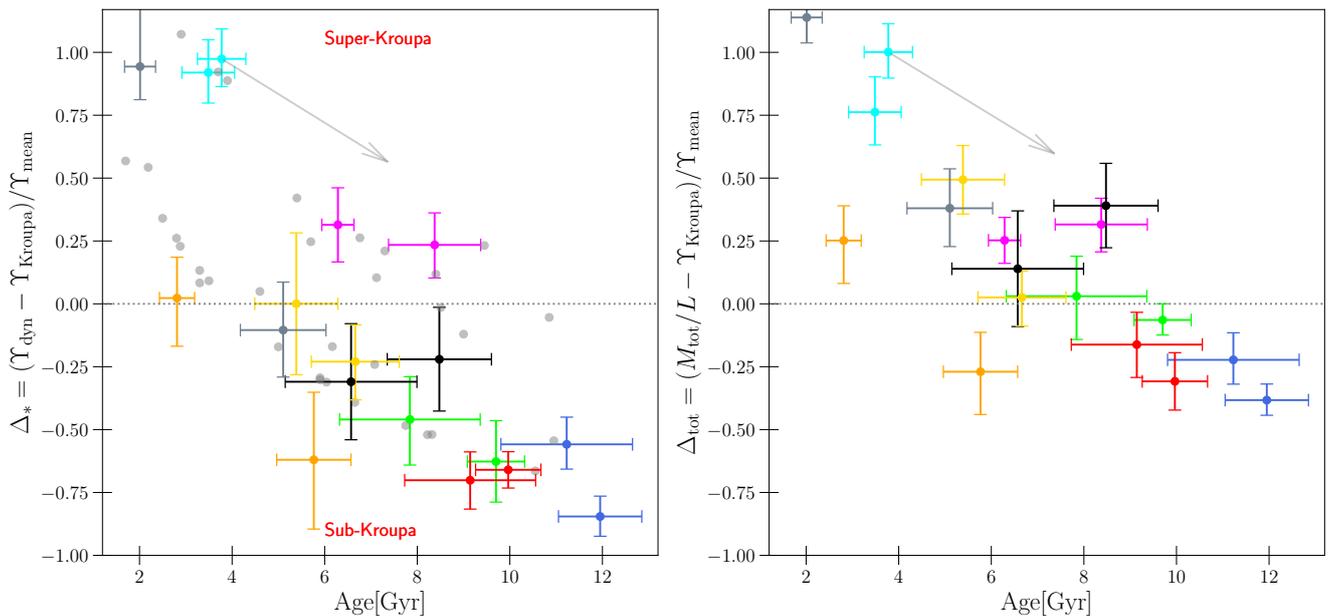


Figure 17. *Left panel:* The relative differences Δ_* of stellar mass-to-light ratios derived from single stellar population analysis Υ_{Kroupa} and dynamical modeling as a function of age of the SSP. *Dots with error-bars:* The differences of the two mass-to-light ratios we derived from the VIRUS-W spectra binned at $r = 2.5''$ and $r = 7.5''$ (Sec. 2.3). *Gray Dots:* The differences when Υ_{Kroupa} values are based on literature stellar population results (cf. App. B). The *dotted, horizontal* line indicates identical population and dynamical results. Assuming the dynamical results are correct, galaxies that lie above this horizontal line would require a super-Kroupa IMF while galaxies below it a sub-Kroupa IMF. *Right panel:* Similar to the left panel, but instead of comparing Υ_{Kroupa} with the *decomposed* stellar component Υ_{dyn} , the population ratio is compared to the *total* dynamically mass-to-light ratio M_{tot}/L , i.e. the dark matter component is included in the mass. Using VCC 2048 as an example the *diagonal arrow* symbolizes how an (unrealistically) large error in the SSP age would shift Δ_* and Δ_{tot} . We measure VCC 2048’s population to be young, but if it was actually twice as old (~ 8 Gyr), Υ_{Kroupa} would be higher and the galaxy should actually be at the arrow’s head if we measured its age correctly.

included. In that case we expect Υ_{dyn} to be completely independent of age, yet the Δ_* would still appear to be anti-correlated with age at least qualitatively. This is because by definition the Υ_{Kroupa} of the population models are increasing with age, which auto-correlates Δ_* and age. The direction of this auto-correlation is indicated by the arrow¹⁶ in Fig. 17.

However, we find the correlation between Δ_* and age we observe can not be driven by uncertainties in the mass *decomposition* alone. This can be seen in the *right* panel of Fig. 17 which displays the differences Δ_{tot} of Υ_{Kroupa} relative to the *total* dynamical mass-to-light ratios M_{tot}/L . The total mass-to-light ratio M_{tot}/L includes the dark matter halo and, thus, reflects the total dynamical mass constraints. Consequently Δ_{tot} does not depend on the merits of the mass decomposition but instead compares the dynamically (required) mass to the

stellar population mass. Two issues are noticed: i) The correlation of the IMF parameter with age persists even if we include the dark matter; ii) Some of the galaxies (VCC 200, VCC 1861) have a negative Δ_{tot} , suggesting that the stellar population models predict a higher (local) mass than what is dynamically inferred. If correct this implies a lighter than Kroupa IMF for these two dEs or, alternatively, that there may be a systematic offset between the two modeling techniques. We will discuss the latter in App. C. The fact that the anti-correlation with age persists even after including the dark matter means that independently of any mass decomposition, the total amount of dynamical mass relative to the expected stellar mass of a Kroupa IMF decreases systematically from younger to older galaxies.

As mentioned above, Δ_* and population age are auto-correlated through Υ_{Kroupa} , hence, one would expect to ‘see’ an anti-correlation of Δ_{tot} with age even if the dynamical mass-to-light ratio are roughly constant for all galaxies. We can remove this implicit dependence of the IMF parameter on age if we only plot the dynamically recovered mass-to-light ratio Υ_{dyn} versus age, which is

¹⁶ The arrow quantifies how a change in SSP age alone would shift Δ_* , i.e. Υ_{dyn} and metallicity are kept the same. The shift is estimated assuming the single population models of Maraston (2005) with a Kroupa IMF.

shown in Fig. 18. In both dynamical mass-to-light ratios (stars only, and total) the anti-correlation remains, although it appears to be weaker than Fig. 17 indicated. In fact, the total mass-to-light ratio seems to be even more strongly anti-correlated than the stellar mass-to-light ratio.

As a sanity check, we have also computed dynamical models composed entirely of stars with a single, radially constant mass-to-light ratio optimised by fitting the kinematics. The mass profiles of these models are fixed by the light distribution. These models are worse fits with much larger AIC_p values. In the sample median they are worse by $\Delta\text{AIC}_p = 50$. Such fits are deemed as unacceptable models (cf. Sec. 4) and due to their inflexible mass profile they often overestimate the mass of the best model at some radius (e.g. in the centre) and, in exchange, underestimate the mass at other radii (e.g. further outside). Still, the overall mass scale of these best mass-follows-light models as shown in Fig. 15 is consistent with the mass-to-light ratios recovered with the more sophisticated dynamical models that included dark matter and stellar gradients. The anti-correlation with age remains even with the mass-follows-light models and all dynamical mass predictions seem to be fairly robust and independent of the specific choice of the dynamical model.

5.2. Are the SSP properties robust?

As mentioned above, the anti-correlation of the IMF parameter with age (Fig. 17) is not obvious to interpret quantitatively because Υ_{Kroupa} , and therefore also Δ_* , are a function of the derived age. If the derived ages and metallicities are erroneous, or even completely random, one would still expect to see an anti-correlation because larger SSP ages always imply larger Υ_{Kroupa} and vice versa. This effect is again quantified by the arrow shown in Fig. 17. If the population of a galaxy that we measure to be young is actually twice as old, then Δ_* and Δ_{tot} would be reduced significantly. While such an unrealistically large error in age quantitatively is still not sufficient to bring all galaxies to a single universal level of Δ_* , it makes it difficult to judge the strength of the correlation by eye.

This is why a comparison of two independent measurements (like Fig. 18) is essential to quantify the strength or existence of such a correlation. Assuming the conditions are comparable for all dEs, e.g. differences in halo formation, environment, etc. are not a major factor in the measured Υ_* differences, then one may expect the dynamical mass-to-light ratios to be a function of the stellar population age. When a galaxy has stopped forming new stars and passively ages, we expect its Υ_*

and M_{tot}/L_z to gradually increase¹⁷. If we assume a universal IMF we can quantify the expected change of the mass-to-light ratio from this aging process by evolving a typical stellar population. This is illustrated for a Kroupa IMF by the dotted line in Fig. 18. In terms of strength this expected mass-to-light ratio change per Gyr is roughly comparable to the trend suggested by the measured M_{tot}/L_z vs age measurements, but its sign is reversed. If we assume a universal IMF and robust dynamical measurements, this implies the actual age trend is roughly opposite to what is measured with the SSP models.

Even assuming the worst case for the SSP models, which would be that they are entirely unconstrained and the derived ages are completely random, it seems strange that the dynamical and population measurement conspired to produce exactly the opposite to what one would expect from the mass-to-light ratio vs age behaviour. For the anti-correlation of the IMF parameter to be meaningless and solely be a result of errors in the SSP properties it would not be sufficient to randomly re-draw ages. Instead it would require the dEs we measured to be very old (~ 12 Gyr) to actually be very young (~ 4 Gyr) and vice versa.

While such a coincidence can not be excluded, given such a small sample, we also have no reason to believe that the SSP properties are entirely random either. The stellar population ages are a direct consequence of the measured $\text{H}\beta$ indices (the strongest age indicator in the VIRUS-W spectra) which is shown in Fig. 19. As expected, the $\text{H}\beta$ indices are anti-correlated with age and have realistic values. This suggests that there is no substantial issue within the SSP modeling itself.

Of course an error could already happen at the level of the index measurement. For example, unmasked emission lines can partially fill up the absorption lines, distorting the corresponding Lick indices, or similarly, problems in the spectral continuum determination or flux calibration could affect the indices. However, we did not find any evidence of emission lines in the spectra. Furthermore the comparison with previously published SSP results shows the same relative trends in age and metallicity we found (Fig. 24), e.g. if we measure one galaxy to be younger and more metal rich than another,

¹⁷ Of course younger galaxies could still have dust/gas left which could boost the mass-to-light ratio. However, none of the dEs in our sample exhibit any signs of dust, and they all have a fairly spatially constant color (Ferrarese et al. 2006) apart from the negligibly small central nuclei (cf. App. E). If there still are small undetectable differences in the dust content, we do not expect them to be able to explain the large variation in mass-to-light ratios across age we find (Fig. 18).

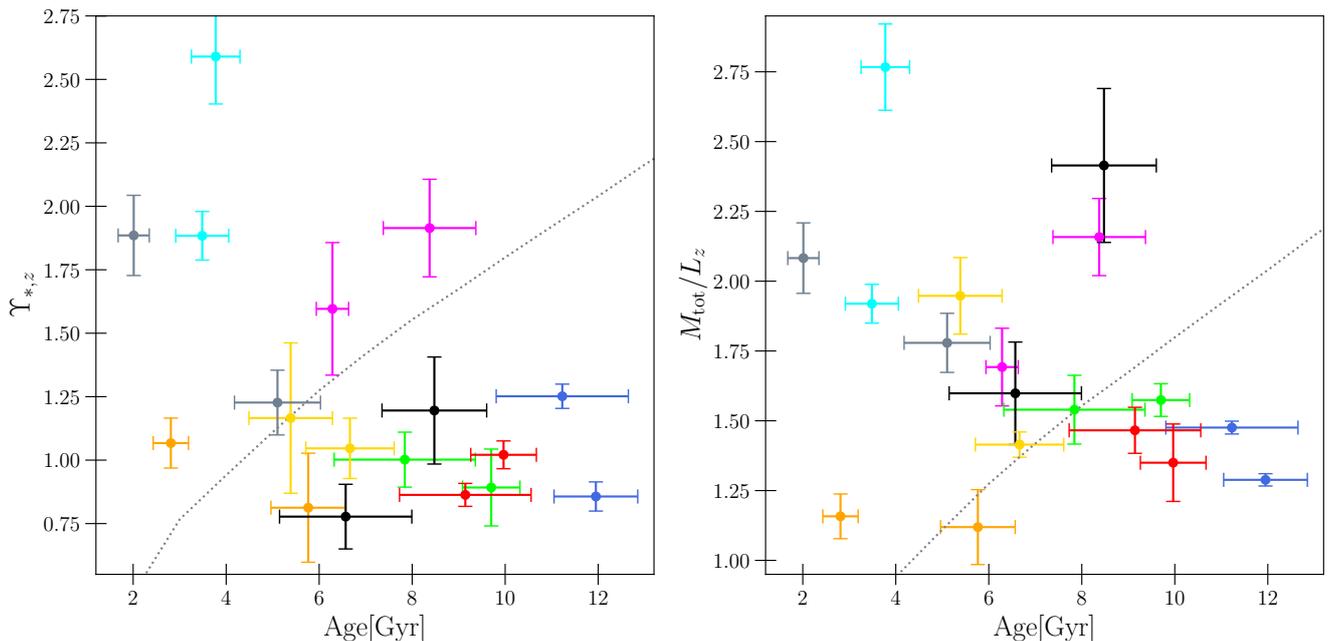


Figure 18. *Left:* The dynamically recovered stellar mass-to-light ratios $\Upsilon_*(r)$ evaluated at radii $r = 2.5''$ and $r = 7.5''$ versus SSP age in the corresponding annuli. *Right:* The same but for the total mass-to-light ratio (incl. dark matter). For this comparison we calibrated all mass-to-light ratios to the z -band. The mass-to-light ratios appear to be anti-correlated with age. If the IMF is assumed to be universal across formation epochs, then this trend is counter-intuitive because one would expect mass-to-light ratio is *positively* correlated with the time passed since star-formation. The dotted line indicates the expected present-day z -band mass-to-light ratio for a single stellar population with a universal Kroupa IMF and a metallicity of $[Z/H] = -0.3$ (Maraston 2005) as a function of its age.

then this is also detected in the other results, even if the absolute differences in age and metallicity may be slightly offset.

In conclusion, for the anti-correlation of the IMF parameter to be solely a result of an uncertainty in the derived SSP ages, a systematic error at the level of the Lick indices present throughout the literature studies would be necessary that is strong enough to be able to change and reverse the derived trends with SSP properties. Or, alternatively, both SSP models and dynamical models have very large uncertainties but coincidentally produced the anti-correlation. A coincidence is not necessarily excluded given the small sample size. To tackle this, a larger number of independent and robust measurements of the *dynamical* mass-to-light ratios will be needed. However, as of now not many dynamical constraints have been published and the existing ones rely on rough virial mass estimates or mass-follows-light models. These rough estimates are often further impacted by the low spectral resolution of the data which biases the inferred velocity dispersions (App. B).

5.3. Star formation history

An assumption that comes along with the SSP modeling is that all stars have formed in a *single* rapid burst

of star formation. However, this may not be reflective of the actual star-formation history (SFH) the dEs have experienced. Two alternatives (or a mixture of both) are conceivable: i) The dEs have a bursty history, i.e. the galaxies are constituted of multiple distinct stellar populations that formed violently in multiple, separate bursts. ii) The dEs had a slow but prolonged star formation over several Gyr before they were quenched.

A bursty SFH can happen if internal or external processes (e.g. mergers, ram-pressure-stripping, re-accretion, ...) are able to continuously trigger and stop multiple star-formation bursts. A plausible scenario is that after a violent initial burst, the resulting supernova feedback expels most of the gas temporarily (but not indefinitely) into the intracluster medium before the galaxy re-accretes the gas again, triggering a secondary star burst (Seo & Ann 2023). Such a process would depend on both environment and total initial mass. For example, a high mass dE in a low-density environment will be able to hold onto more of its gas without expelling it indefinitely and, thus, is able to trigger multiple bursts.

A slow and gradual SFH on the other hand can happen if the galaxy has a continuous supply of gas (e.g. through wet minor mergers or further gas accretion). Again high-mass dEs that inhabited low-density envi-

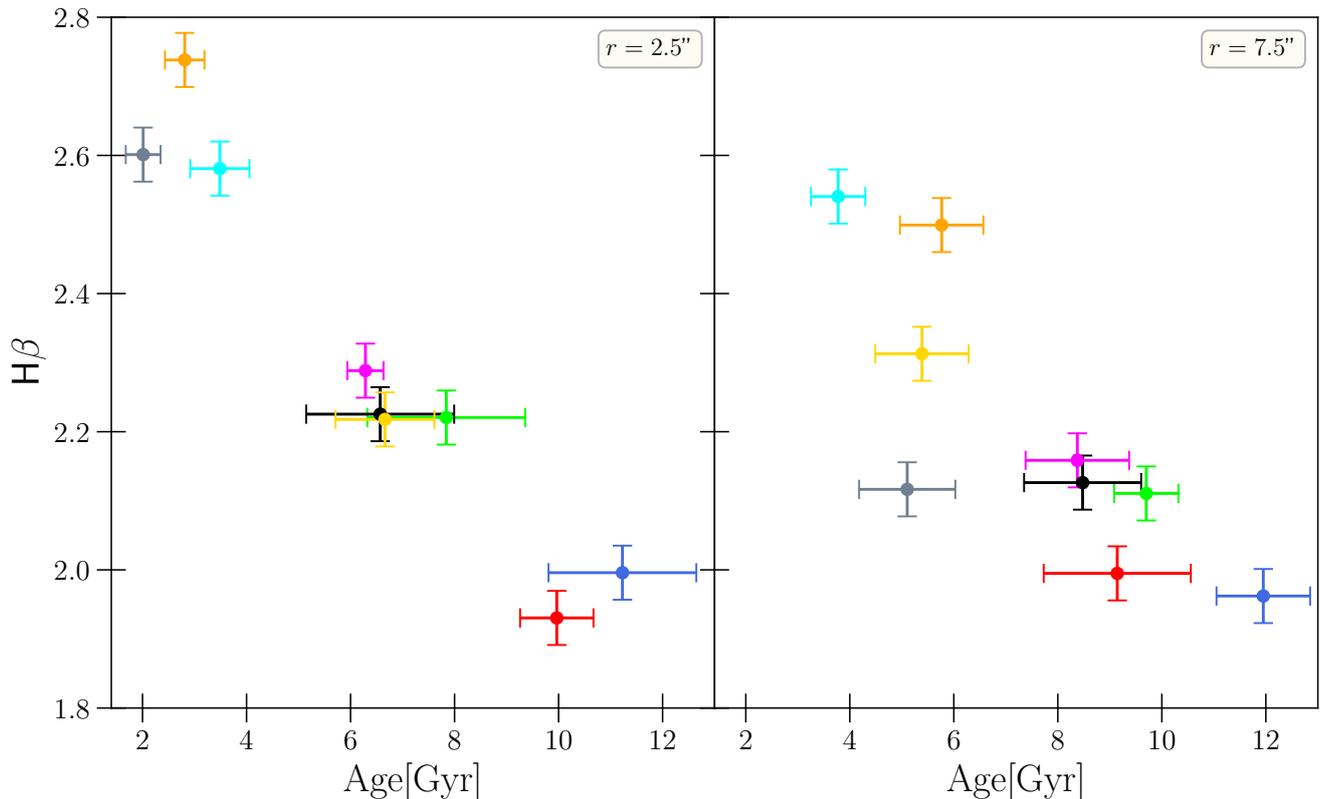


Figure 19. The $H\beta$ index vs. the SSP age derived from the two spectra at $r = 2.5''$ (left) and $r = 7.5''$ (right). The measurement of the strong age-indicator $H\beta$ appears to be a strong predictor of the age, which suggests the age we derived is a direct result of the measured spectra and not a random modeling result.

ronments in their past and, thus, gravitationally dominated their local environment are likely candidates for such a prolonged SFH. The galaxies would have continuously formed stars during their LTG progenitor phase only stopping when quenched and transformed to a dE (e.g. as they experience their first infall into the cluster).

In either case (bursty or continuous) the extended SFH could lead to the anti-correlation of the dynamically measured mass-to-light ratios with age (Fig. 18). Supposing all dEs have formed in the same epoch (e.g. 12 Gyrs ago) but depending on their total mass and environment, some of them (the ones we find to be younger and metal-rich) were able to have an extended SFH. Then the apparently young and metal-rich galaxies could have significant amounts of older stellar populations with a higher mass-to-light ratio, which may be overshadowed by the youngest component from the last star-formation period. Moreover, they could host far more stellar remnants (from long passed SF periods) than the SSP models would predict. In that case the mass bound in stars would be higher than the luminosity suggests and the actual mass-to-light ratio would

be higher than that of the most recently formed sub-population which might dominate the SSP model.

Several of our findings support this SFH explanation for the Δ_* -age anti-correlation. Firstly, the younger galaxies appear to have almost solar-like metallicity (Fig. 24), suggesting the youngest population has been chemically enriched by the past SFH. Secondly, the two youngest galaxies are also found in the lowest density environments near the cluster’s virial radius (Fig. 8), which may imply they have only recently been quenched. Thirdly and lastly, the dEs generally have low, near solar-like $[Mg/Fe]$ ratios when compared to the higher ratios of more massive ETGs¹⁸, which implies the SFH of dEs were more gradual. This is not new and numerous studies of several abundance ratios have found that dEs have a chemical composition that indicate a prolonged SFH (e.g. Geha et al. 2003; Michielsen et al. 2008; Şen et al. 2018; Romero-Gómez et al. 2023b).

¹⁸ A lower $[Mg/Fe]$ is an indication of a more extended SFH because Mg (as an α element) is made predominately by supernova of type II whereas Fe builds up more gradually via type Ia.

If the SSP ages we find are more representative of the epoch of last star-formation (i.e. of the time they were being quenched) and less so of the galaxy’s time of formation, then we may expect that the [Mg/Fe] ratios are correlated with the SSP ages. Indeed we find signs of such a correlation as illustrated in Fig. 20. Especially in the center ($r = 2.5''$), the abundance ratios are strongly correlated with age (more so than with environment, see Fig. 8).

If the dEs indeed have experienced a complex star-formation history over several Gyr, then one may be able to unravel the Δ_* -age anti-correlation by modeling them with multiple populations (if one can get the degeneracies under control). As mentioned in App. B, Ryś et al. (2015) have tested the simplest form of extended star formation (2-burst models) on some of our sample galaxies. Unfortunately, for most of those dEs they often had trouble finding a solution for the older population, in which case they fixed its age to 12 Gyr to find a solution. Still, qualitatively, their results are fairly supportive of a complex, extended SFH. For example, while 70 to 80% of the light of VCC 2048 (SSP age of ~ 3 -4 Gyr) comes from a young population (2 to 3 Gyr), about half the mass is currently found in the second older 12 Gyr population. While we can expect the IMF parameter of the dEs with old SSP ages (~ 12 Gyr) are robust (the *single* population assumption is correct), this is less clear for the dEs with younger SSP ages. In a scenario like the above for VCC 2048 where half the mass is bound in an older ‘hidden’ population, a single-population-model would result in an underestimation of Υ_{Kroupa} and therefore an increase of Δ_* . As such the observed Δ_* -age anti-correlation could be an artifact originating from not accounting for the extended SFHs of dEs with younger ages. However, quantitatively it seems implausible that the SSP assumption alone could be responsible for the variation in the IMF parameter: to bring the very high Δ_* of the youngest dEs (e.g. VCC 2048) down to the sub-Kroupa level of the oldest (e.g. VCC 200) their Υ_{Kroupa} would have to be up to 5.77 times higher than what we estimated with the SSP models. Nonetheless we may expect that the assumption of a single population could artificially increase the significance of the Δ_* -age correlation. A robust modeling of the (potentially) complex SFH of dEs will be needed in the future.

It is likely that galaxies in the mass regime of our dE sample are particularly prone to display a complex extended SFH. More massive galaxies quench quickly because of feedback from active galactic nuclei. Less massive galaxies also quench quickly because they are much more vulnerable to rapid quenching by their envi-

ronment or supernova feedback. However, in the intermediate regime of the dEs there may be a ‘sweetspot’ at which a galaxy is just about massive enough to hold onto its gas reservoir for a prolonged period without ejecting it due to violent internal feedback. For example, while for the less massive galaxies gas could be removed indefinitely and rapidly in a *single* quenching event, the slightly more massive dEs could be able to reaccrete some the ejected gas and rejuvenate their star formation. The results of Romero-Gómez et al. (2023a) indeed suggest that such a maximum-SFH sweetspot is somewhere between 10^8 and $10^9 M_\odot$. They analysed the SFHs of 3 different galaxy samples (Local Group dSphs, Fornax dEs, massive ETGs) and found that the dEs formed their stellar mass more slowly than the galaxies in the other two samples. Surveys that study the mass regime between $10^6 - 10^{10} M_\odot$ using a *single* cohesive sample/analysis may be able to confirm the existence and location of this sweetspot in the future.

5.4. IMF variation with formation time

If an extended SFH is not sufficient to explain the correlations of Δ_* with age and metallicity quantitatively and the mass-to-light ratios from dynamical and SSP modeling are indeed robust, the alternative physical explanation is that the assumption of a universal Kroupa IMF is wrong and instead the true IMF changes over time and metal content. To be consistent with the dynamical masses, the galaxies with $\Delta_* > 0$ would require a *Super*-Kroupa IMF, while galaxies with $\Delta_* < 0$ require a *Sub*-Kroupa IMF. In that case the correlation suggests that the IMF changed from Sub-Kroupa IMF in the early Universe to a Super-Kroupa IMF in the present, while crossing the intermediate Kroupa IMF at around 4-8 Gyr ago. Our results are not the first indication that masses derived from SSP-modeling with an assumed universal IMF maybe be overestimated for older populations. For example, the anti-correlation observed in ETGs between central dark matter fraction and SSP age may well be explained by a lighter IMF of older populations (cf. Napolitano et al. 2010; Tortora et al. 2014). This is also not the first time a variation of the IMF with time and metallicity content is proposed (e.g. van Dokkum 2008; Li et al. 2023), with the theory being that at early formation epochs the IMF is more ‘bottom-light’, or analogously more ‘top-heavy’, than present-day IMFs. Such IMFs may be caused by the on average higher temperatures of the star forming clouds in the early Universe (e.g. due to CMB heating) or more effective stellar feedback in the early, low metallicity environments (Larson 2005; Chon et al. 2022, 2024).

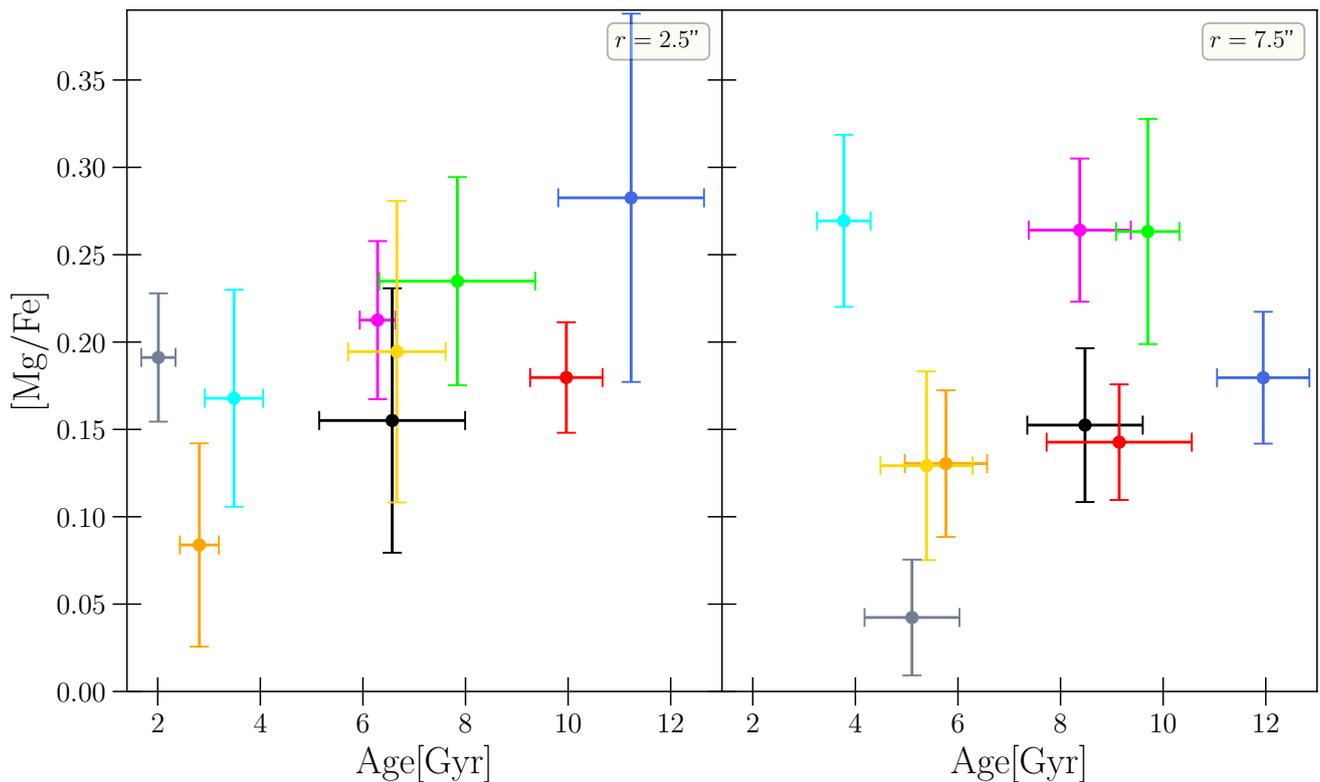


Figure 20. The abundance ratio $[\text{Mg}/\text{Fe}]$ (a proxy for extended star formation) vs the single population age. Especially in their center ($r = 2.5''$), dEs that are young also have lower solar-like $[\text{Mg}/\text{Fe}]$, which suggests they had a more extended star formation history than older ones which (presumably) were quenched earlier.

Regardless, the large change in Δ_* we observe across our sample galaxies does require a significant change of the IMF during the lifetime of the Universe. At this point the variation (or universality) of the IMF with time is still a hotly debated issue (for a review see Bastian et al. 2010). While our dE sample does suggest a potential variation of the IMF across formation epoch, the evidence across the literature is still conflicting depending on the IMF measurement probe that is being used. For example, while our dwarf models suggest a *negative* correlation of the IMF parameter with age, observations of ‘relic’ galaxies (which are assumed to be probes of conditions in the early Universe) suggest the opposite trend: a *positive* correlation with age (Martín-Navarro et al. 2023). (Parikh et al. 2018) who studied the IMF of ‘ordinary’ ETGs with masses $\log(M_*/M_\odot) \sim 10 - 11$ find a similar, but stronger, metallicity trend as we do but their trend with age is again the opposite. If on the other hand we analyse the IMF parameters and SSP ages of the ATLAS^{3D}-survey (Cappellari et al. 2011) we find no strong correlation with age in either direction. Only the few young ATLAS^{3D} galaxies with $\sigma < 100 \text{ km s}^{-1}$ tentatively show signs of Δ_* -age anti-correlation.

In conclusion, while some of the above discussed potential causes for the Δ_* -age correlation we found appear more plausible than others, we believe a superposition of the discussed issues could be the most likely explanation. To confirm the veracity of this correlation, a larger sample will have to be investigated that ideally includes a larger variety in total mass and environments (e.g. field galaxies). At the current level of accuracy we do not believe we can employ the SSP results to complement or help improve the dynamical models of our dE sample, e.g. by using them as additional constraints to help decompose baryonic and dark matter (VW-II). Neither do we deem the SSP results to be robust enough to corroborate the veracity of the dynamical mass decomposition. Future simultaneous population and dynamical analysis applied to the same data sets, such as done in this study, may be very conducive to test current modeling assumptions and confirm or deny our findings regarding the IMF and/or a prolonged SFH.

A problem with investigating a potential non-universality of the IMF due to age (or metallicity) is that the single stellar population estimates are most sensitive to the epoch when a galaxy has *stopped* forming stars than to the epoch when the galaxy has actually formed.

Complementary to more sophisticated population models which may be able to account for complex SFH one may also look for probes of the galaxy formation epoch that are independent of stellar activity. Studying galaxies’ dark matter halo densities (VW–II) could be such a way to gauge the epoch of gravitational assembly independently and to discriminate whether the dEs have assembled in different epochs (implying a non-universality of the IMF with age) or all at a similar, early epoch (implying varying degrees of SFH and the IMF trend with age to be an artifact).

6. SUMMARY AND CONCLUSIONS

We have presented new kinematic data for a sample of 9 Virgo-Cluster dEs obtained with the high-resolution IFU-spectrograph VIRUS-W. It is the first 2D kinematic study of these objects with such a high spectral resolution which allows the recovery of velocity dispersions reliably down to $\sigma \sim 15 \text{ km s}^{-1}$ and out to approximately $1 r_e$. Previous observations with lower spectral resolution tend to yield dispersions that were biased high by up to 20–50% (App. B). We also provide the first spatially resolved measurements of the higher order non-gaussian moments of the LOSVDs for these galaxies. The dEs exhibit diverse but systematic kinematic signatures in all Gauss–Hermite moments up to h_4 . We find dEs with strong central dispersion drops as well as flat and slightly decreasing dispersion profiles. Many of the galaxies follow the same $v - h_3$ anti-correlation known from more massive galaxies. Some of the dEs also has a central peak in $h_4 \sim +0.1$, whereas the profile decreases radially to ~ 0 which is often associated with a central dispersion drop.

Mild dynamical mass-to-light ratio gradients.

We have used the spatially resolved LOSVDs to construct orbit-superposition models which allowed us to dynamically constrain their 3D intrinsic mass and kinematic structure. This study is the first attempt to dynamically constrain stellar mass-to-light ratio *gradients* on the scales of dEs. The gradients we recover are generally low to moderate but we find a strong positive correlation of the stellar mass-to-light ratio gradients with the observed line-of-sight velocity dispersion profiles. Galaxies that feature a radially increasing σ also increase in their stellar mass-to-light ratio with radius and vice versa, while dEs with a flat dispersion profile are also flat in their mass-to-light ratio. Averaged over the whole sample, the gradients are distributed around zero.

SSP gradients. We also binned the spectra in two annuli to perform a single stellar population analysis with the goal to derive age, metallicity and abundance

ratios. Our SSP results are in broad agreement with most of the existing literature, but the scatter and measurement error are considerable. Unlike the case for the LOSVDs, we have no reason to believe that our SSP results are significantly more reliable than the already existing measurements. We find little to no correlation between SSP results and cluster environment. Combining the SSP analysis with the dynamical results we do *not* find any strong evidence of a spatial variation of the IMF as found for the bulges of massive ETGs (Mehrgan et al. 2024; Parikh et al. 2024). On average dEs are consistent with a Kroupa-like IMF. Taken together, both SSP gradients and dynamical models paint a spatially homogeneous picture of the luminous matter. Apart from some outliers for which the interaction with the intra-cluster medium has presumably rejuvenated recently some central star formation, the stellar population properties and stellar mass-to-light ratios of dEs change at most only moderately. This suggests that the bulk of a dE’s stars (within the apertures we investigated) has formed in parallel from the same IMF and dEs were quenched all at once.

Anti-correlation between M/L and age. While the stellar changes within a single galaxy might be small, we find a much larger heterogeneity across the different sample dEs. Dwarf ellipticals display a greater diversity in age compared to the generally much older giant ETGs, with dEs SSP ages ranging from 2 to 12 Gyr. We find that their mass-to-light ratios are anti-correlated with this single population age, i.e. the younger the galaxy is, the more the dynamical mass-to-light ratio exceeds what is expected for a Kroupa IMF. This either suggests a dependence of the IMF on formation epoch, or alternatively, a more complex, prolonged star formation history. The former scenario suggests the Virgo dEs are constantly being produced in different formation epochs and environments, while the latter suggests all dEs have formed at roughly the same time early in the Universe, but the ones found to be young have had prolonged, complex star formation history whereas older galaxies were quenched early on. Compared to ‘ordinary’ ETGs, little is known about the IMF of dwarf ellipticals yet, but our results suggest that future IMF studies will require more sophisticated population models than what was used for most ‘ordinary’ ETG studies. To establish conclusively whether the IMF varies with age or not will require sophisticated extended SFH modeling.

Suppressed angular momentum. In terms of their *projected* angular momentum, the majority of dEs are either classified as intermediate rotators or slow rotators. The result stands in opposition to the angular

momentum parameter of more massive ETGs. A plentitude of large samples, which study ETGs in the high mass ranges $\sim 10^{10} - 10^{12} M_{\odot}$, have identified a transition in the kinematic structure of ETGs around $10^{11} M_{\odot}$ (e.g. Emsellem et al. 2011; Jin et al. 2020; Santucci et al. 2023) where angular momentum parameter and orbit structure change dramatically. For the most massive ETGs, the angular momentum parameter is low, more stars occupy more radial orbits, and they form tangentially anisotropic cores due to black hole scouring. In contrast ‘intermediate-mass’ ETGs ($\leq 10^{11} M_{\odot}$) can often have more tangential orbit contributions and higher angular momentum. However, the results for the dEs suggest a *second* point of change in the kinematic structure that occurs in the ETG sequence at even lower masses around $M_{*} \sim 10^{9.5} M_{\odot}$, at which point the amount of ordered motion is reduced again. This trichotomy in the ETG mass sequence appears to be traced by a corresponding behaviour in the total specific angular momentum $j = J/M$, such that dEs have significantly less angular momentum than expected from (dark matter only) structure formation models. The correlation of the angular momentum of dEs with their environment suggests that external influences play a role in this reduction. The galaxies become more susceptible to tidal perturbations and ram-pressure stripping as their total mass decreases. However, even the dEs in our sample that are at the cluster’s virial radius appear to be only moderately rotating, which could suggest that the environment can not be the sole driver of the momentum reduction. *Internally* induced heating processes could also play a significant role, and similar to the external processes, their impact can be expected to increase as the potential well of the galaxies becomes shallower. Together both external and internal influences may play an important role in shaping the orbital structure of dEs, but instead the suppressed angular momentum could also be more so a result of their distinct gravitational assembly rather than the result of those secular processes. This will be explored further in VW–II. To investigate these different processes, future studies that include genuine field dEs far outside any cluster or group may be helpful to break this degeneracy.

A 3D orbit structure aligned with its star-forming progenitor. The above ‘trichotomy’ in the angular momentum of ETGs is also echoed in the *intrinsic* orbital structure. The dEs behave distinct from both intermediate mass ETGs and the most massive ETGs. In spherical coordinates the intrinsic anisotropy parameter β is close to isotropic or mildly radial. In contrast, more massive ETGs typically exhibit an anisotropy structure that varies more strongly with radius and is

far off from isotropy. Intermediate mass ETGs exhibit a range of different orbital structures and can be both significantly radial as well as tangential. The more massive cored ETGs are more homogeneous to each other, but have tangentially anisotropic cores due to black hole scouring. If we analyse the dEs in cylindrical coordinates, we find that intrinsically flattened dEs have an increased kinetic energy in their equatorial plane despite their aforementioned suppressed angular momentum. Flattened dEs in our sample tend to have a higher σ_{ϕ} in the equatorial plane rather than enhanced radial motions which contributes significantly to their flattening by the (low) angular momentum. This orbit structure makes (dry) mergers very unlikely, and it may be interpreted as a (partially) intact relic of their star-forming progenitors. The low energy perpendicular to the disk again poses the question of how efficient external/internal heating mechanism really are, and whether the low angular momentum is more so an expression of dE assembly (VW–II).

All in all, dEs appear to be *spatially* homogeneous in their stellar structure within the scales we investigated. Any stellar sub-structure must be too weak to be detectable on the macro level of the dynamical and population models. Still, detailed photometry studies (e.g. Barazza et al. 2002) suggest dEs exhibit, even if only faint, substructure (see also discussion App. E). In our sample galaxies most notably are the unresolved blue nuclei which seem to be more or less present in any dE, even if initially classified as a non-nucleated dE. Furthermore, some dEs may have weak embedded disks as suggested by faint face-on spiral arms or disk-like isophotes. Still, currently we do not find compelling evidence for the need of 2-component model structure (e.g. disk, bulge, bars,...) on a macro level. Perhaps future studies with a higher resolving power and S/N may be able to use the detailed photometric substructure in their analysis.

The homogeneous *spatial* structure we observe in the dEs contrasts with significant heterogeneity across the different sample galaxies. The dE population seems to be very different in their stellar mass and kinematic structure when compared to the rest of the ETG sequence. These differences may be explained by the very different evolutionary channel the dEs follow when compared to the more massive ETGs: For the latter a history of mergers and continued accretion of material has played the dominating role in shaping the orbit and mass-to-light ratio structure. Whereas for the dEs internal feedback process and external influences by the environment may have changed them significantly. This may have also left an imprint in the dark matter halos

and shape of the mass distribution which we will investigate in VW–II.

Facilities: Smith, HST

Software: astropy (Astropy Collaboration et al. 2022)

ACKNOWLEDGEMENTS

We thank the anonymous referee for comments and suggestions that helped improving the paper. This work is based on observations obtained with the Harlan J. Smith Telescope at the McDonald Observatory, Texas. Computing has been carried out on the COBRA and RAVEN HPC systems at the Max Planck Computing and Data Facility (MPCDF), Germany. We also made frequent use of the NASA/IPAC Extragalactic Database (NED), operated by the Jet Propulsion Laboratory and the California Institute of Technology, NASA’s Astrophysics Data System bibliographic services, and the HyperLeda database (Paturel et al. 2003).

APPENDIX

A. EXAMPLE FOR A TYPICAL LOSVD RECOVERY

Fig. 21 shows an example for a VIRUS-W spectrum fit used to extract the kinematical and stellar population properties as is outlined in Sec. 2.2. A comparison of the LOSVD recovery with a Gauss–Hermite model vs a non-parametric model description of the LOSVD is shown in Fig. 22. In this case the Gauss–Hermite model is preferred as it achieves a slightly smaller AIC_p . Still, generally both LOSVD model choices are fairly consistent with each other for a given Voronoi bin.

B. COMPARISON TO EXISTING KINEMATIC AND STELLAR POPULATION STUDIES

In the last two decades, almost all of the galaxies in our sample were the subject of kinematic and stellar population studies already. In the following we discuss and compare our kinematic and population results with the studies we are aware of, discuss the significance of our results and highlight the need for a high resolving power. The published studies we compare our results to usually include other dEs as well, but we will focus only on those galaxies that are also part of our own sample. We plot the age, metallicity, rotation velocity and velocity dispersion from previous measurements together with our results in Fig. 23 and Fig. 24 (for those studies where we were able to extract the respective data in a consistent way).

As part of a larger sample of early-type galaxies with various sizes, the kinematics of VCC 543, VCC 856, and VCC 2048 were studied by Simien & Prugniel (2002) using a long-slit spectrograph, which achieved a resolving power of $R = 5050$, or an instrumental dispersion that is just below, or at least at a similar level to, the minimum velocity dispersion we measured for these galaxies. They measured the velocity and dispersion of the spectra out to $\sim 20\text{--}25''$, finding velocity curves consistent with ours, although the scatter is considerable. The dispersions of VCC 543 and VCC 856, however, show significant, qualitative and quantitative disagreements: they scatter a lot, show no clear radial gradient and are generally much higher than the values we find for VCC 543 and VCC 856. For VCC 2048 the dispersion profile agrees qualitatively with ours showing a dispersion drop in the center (albeit shallower). However, again the dispersions are biased high on average. The discrepancies are likely caused by their larger instrumental dispersion which is close to the velocity dispersions of these galaxies. An additional factor could be that Simien & Prugniel (2002) did not include higher order deviations from Gaussian LOSVDs in their fits which could artificially broaden the LOSVDs.

The spectral resolution was even lower in the studies of Caldwell et al. (2003), van Zee et al. (2004), Michielsen et al. (2008) and Paudel et al. (2010) ($R \sim 1500$, $R \sim 2200$, $\sigma_{\text{instr}} \sim 170 \text{ km s}^{-1}$ and $\sigma_{\text{instr}} \sim 280 \text{ km s}^{-1}$, respectively). Accordingly, when stated, the respective velocity dispersions measured in these studies are again higher than ours. The dispersions of van Zee et al. (2004) are all offset by more than 10 km s^{-1} while those of Caldwell et al. (2003) are

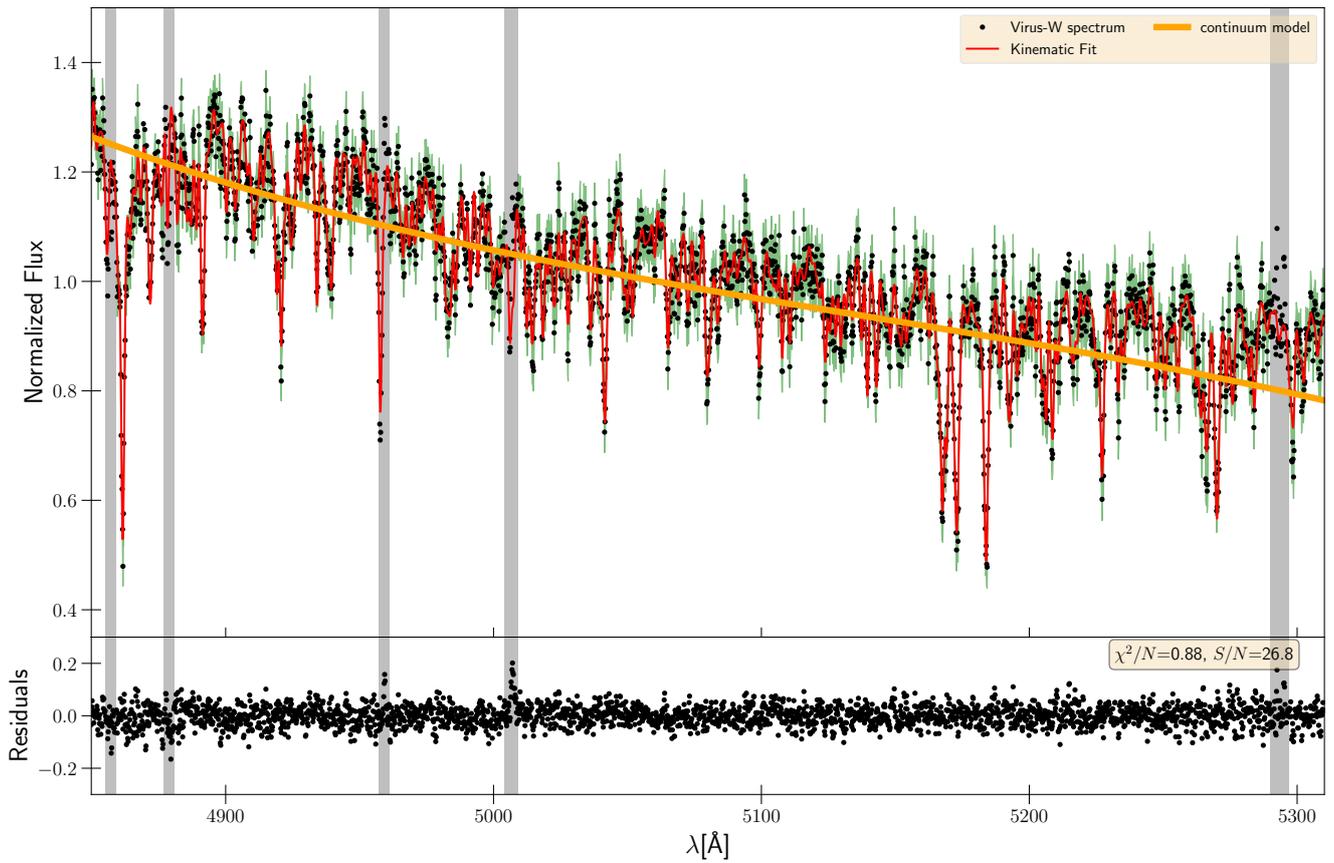


Figure 21. *Top:* Example of a typical Voronoi binned spectrum (one of the bins of VCC 200 within the central $5''$). *Black:* The observed VIRUS-W spectrum. The *green* band around the data indicates the 1σ flux uncertainty. *Red:* The corresponding Fit model. *Orange:* The continuum modelled by a 2nd order multiplicative polynomial. The *gray* regions indicate the parts which were masked before the fit. *Bottom panel:* The Residuals between Fit and Data.

even outside the plot range of Fig. 23. Michielsen et al. (2008) and Paudel et al. (2010) only stated stellar population properties and no kinematics. Caldwell et al. (2003) discussed both population properties and kinematics. The age and metallicity Caldwell et al. (2003) determined for VCC 2048 agree with our results within the one sigma error while their age for VCC 856 is significantly lower. However, they did not find old ages for any of the low σ -galaxies in their sample. The SSP results of Michielsen et al. (2008) are based on a slightly higher resolution and seem to agree better with our results. On the other hand Paudel et al. (2010) found systematically younger and more metal rich populations than we (and other studies) do.

Chilingarian (2009) re-analyzed the data of Simien & Prugniel (2002) and van Zee et al. (2004) using a full spectral fitting technique instead of Lick indices. The dispersions of VCC 543, VCC 856, VCC 1261 are relatively consistent with those older studies in that their scatter is high and the average dispersion is significantly larger than what we find. Only VCC 2048 does not appear to systematically offset high, and in fact it is the only example where a significant portion of the literature dispersions fall *below* our values. While about half of the spatial bins of VCC 2048 fall close to the dispersions we derived, the other half scatters to the low side. The fact that some of these bins have such a low dispersion is peculiar and seems inconsistent with the dispersions from Simien & Prugniel (2002) which were derived from the same raw data. The population properties Chilingarian (2009) derived were spatially resolved by separately analysing the co-added spectra in *two* different radial bins along the slit: A circum-nuclear bin (excluding the central blue nuclei) and an outer bin (near or just within $\sim 1 r_e$). In Fig. 24 we plot both, the values of the circum-nuclear and the outer region. The population properties agree well, and in fact, even the radial behaviour is (at least qualitatively) consistent with our findings.

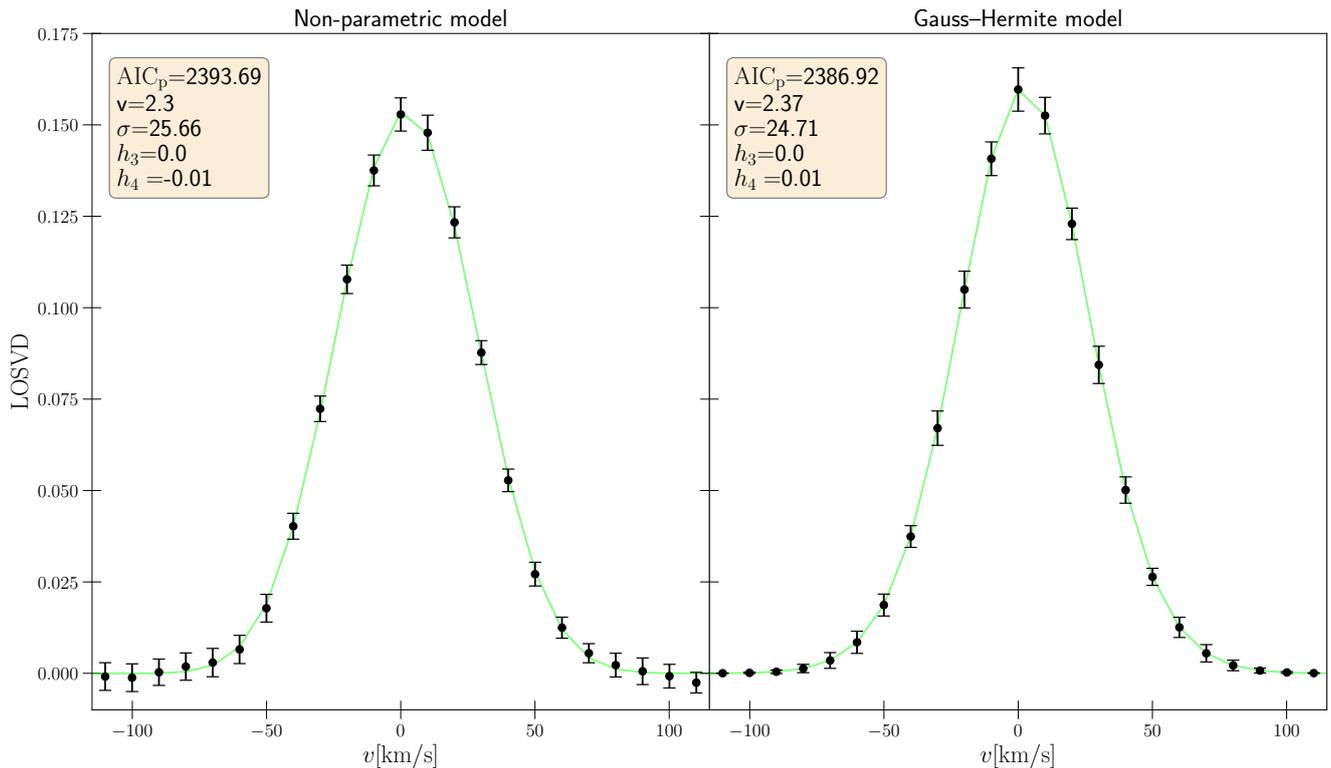


Figure 22. Comparison of an optimized LOSVD model parametrized by a Gauss-Hermite series (*left*) and a non-parametric description (*right*) for the same Voronoi Bin of VCC 200 ($S/N = 26$). Both are broadly consistent with one another. In the case of this specific Voronoi Bin, the Gauss-Hermite representation is slightly preferred by AIC_p . The error suppression by a Gauss-Hermite parametrization at high velocities is substantial which is why we readjusted the errors as described in Sec. 2.2.

Toloba et al. (2014) analysed the rotation, dispersion and population properties of a large dE sample using optical long-slit spectroscopy from 3 different telescopes with the spectral resolution ranging from $R \sim 1900$ to 3300. As is the case with most previous studies, the velocity curves are fairly consistent with our findings. However, unlike many other low spectral resolution studies, they also found quantitatively consistent dispersions for some (but not all) of the low- σ galaxies (specifically VCC 308, VCC 543, VCC 1861). This could be because they aimed for a very high S/N for the purposes of a better dispersion recovery. High S/N can counteract low resolution effects as shown in the recovery simulations of Toloba et al. (2011) and Eftekhari et al. (2022). We plot the ages and metallicity they derived for the summed spectra within $1 r_e$ in Fig. 24 (some of their spectra include more Lick indices than ours, such as the age-sensitive $H\alpha$ line; e.g. for VCC 308, VCC 543, VCC 1528, VCC 1861 they include this index). We excluded VCC 856 from our plot as the age they derived is at the boundary of their grid (14.1 Gyr). Şen et al. (2018) reanalyzed the data of Toloba et al. (2014) with a special focus on different abundance ratios in dEs. For this purpose they derived ages and metallicities for the very central spectra ($r < r_e/8$) from a combination of 23 Lick indices. Overall their results are in good agreement with Toloba et al. (2014) and with ours, not only in age but also in metallicity (Fig. 24).

Geha et al. (2003) is the only comparison study with better data in terms of resolution ($R \sim 13000$) and spectral coverage (3900Å-11000Å) (highlighted in green in Fig. 23). Their dispersions for VCC 543, VCC 856, and VCC 1261 are quantitatively much more consistent with ours. We suspect that minor differences could be caused by the fact that only Gaussian LOSVDs are fitted and because they only used K-type stellar templates.

With the advent of IFUs, 2D kinematic studies of dwarf galaxies have become more prevalent. In a series of papers (Ryś et al. 2013, 2015) were the first to analyze the kinematics and stellar population of the galaxies in our sample using an IFU: SAURON in its low resolution mode ($R \sim 1300$ within the spectral range of 4760Å-5300Å). The velocity maps are fairly consistent in that VCC 1261 and VCC 1861 are essentially non-rotators, while VCC 308 and VCC 2048 show intermediate to strong rotation respectively. Their dispersions on the other hand scatter significantly between neighbouring Voronoi bins with differences of up to 100 km s^{-1} . Therefore in Fig. 23 we only show their radially

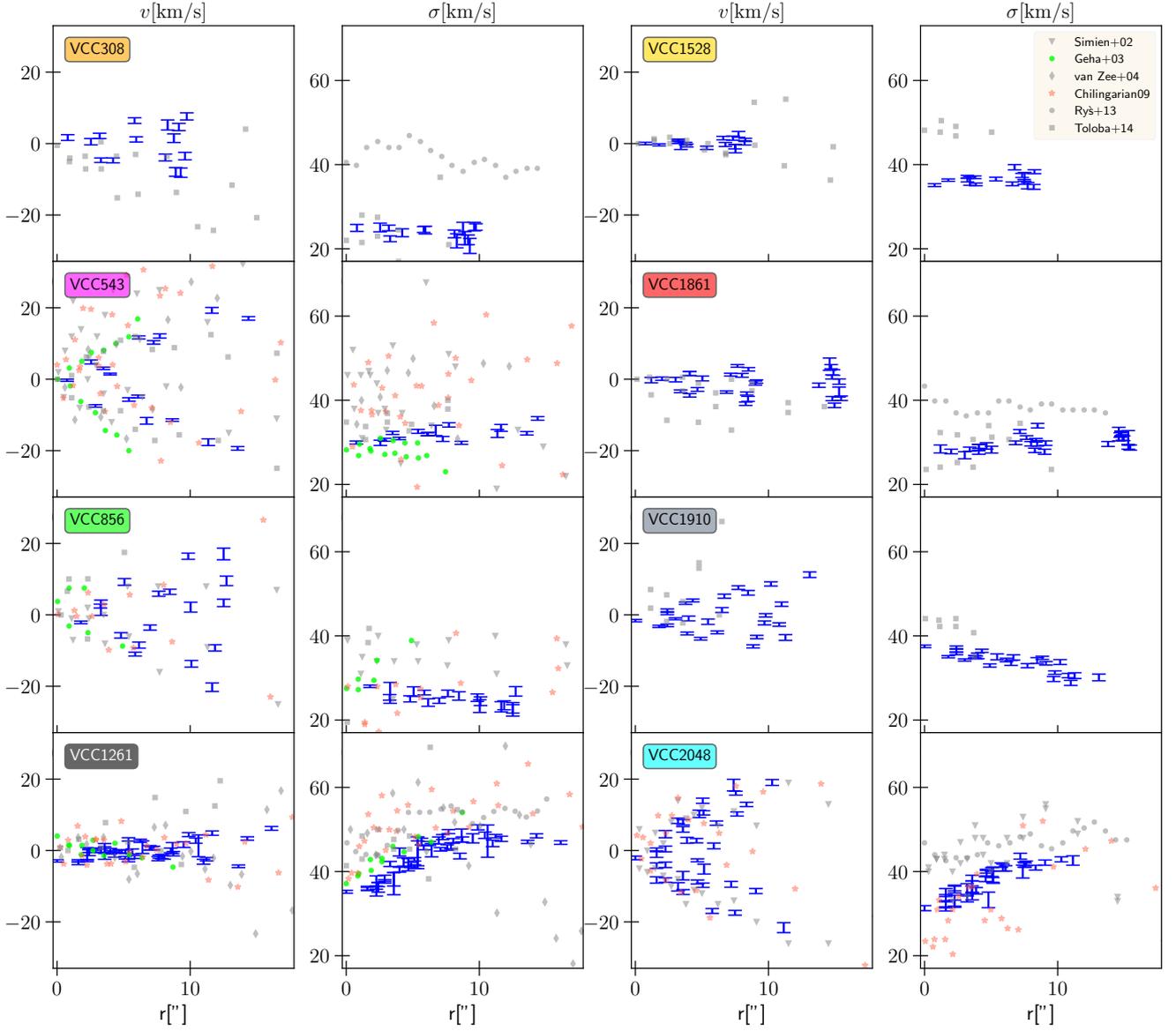


Figure 23. Literature comparison for the mean (line-of-sight) velocity and dispersion. In this paper we differentiate between sample galaxies using the color-coding indicated by the background color of the label here. There is no published data available for VCC 200 so we excluded it from this plot. The color-coding that is used for VCC 200 in all other Figures is *dark blue*. To avoid overcrowding we do not show errorbars for the literature. For the studies where we did not find any raw data in a table format we reverse engineered the approximate values from their Figures. A few literature data points are not displayed here because they lie outside the plot boundaries. *Blue with errorbars*: Our kinematic data (see also Fig. 3). *Gray triangles*: Data from Simien & Prugniel (2002). *Green dots*: Data from Geha et al. (2003). *Gray Diamonds*: Data from van Zee et al. (2004). *Red stars*: Data from Chilingarian (2009). *Gray dots*: Dispersion from the IFU study of Rys et al. (2013). Since they have ~ 100 Voronoi bins we only plot the dispersion averaged within ellipticals bins as shown in Fig.5 of Rys et al. (2013). This reduces their high bin-to-bin scatter in the dispersion significantly, yet the higher dispersion remains. Their mean velocities were not displayed in the same manner which is why they are not shown here. *Gray squares*: Data from Toloba et al. (2014).

averaged dispersions which reduces the bin-to-bin scatter, but still the dispersions seem systematically high. They did not find dispersions below $\sim 40 \text{ km s}^{-1}$. We excluded their ages and metallicities from Fig. 24 because they used models with 2 distinct stellar populations, i.e. they are not directly comparable to single population models. Still,

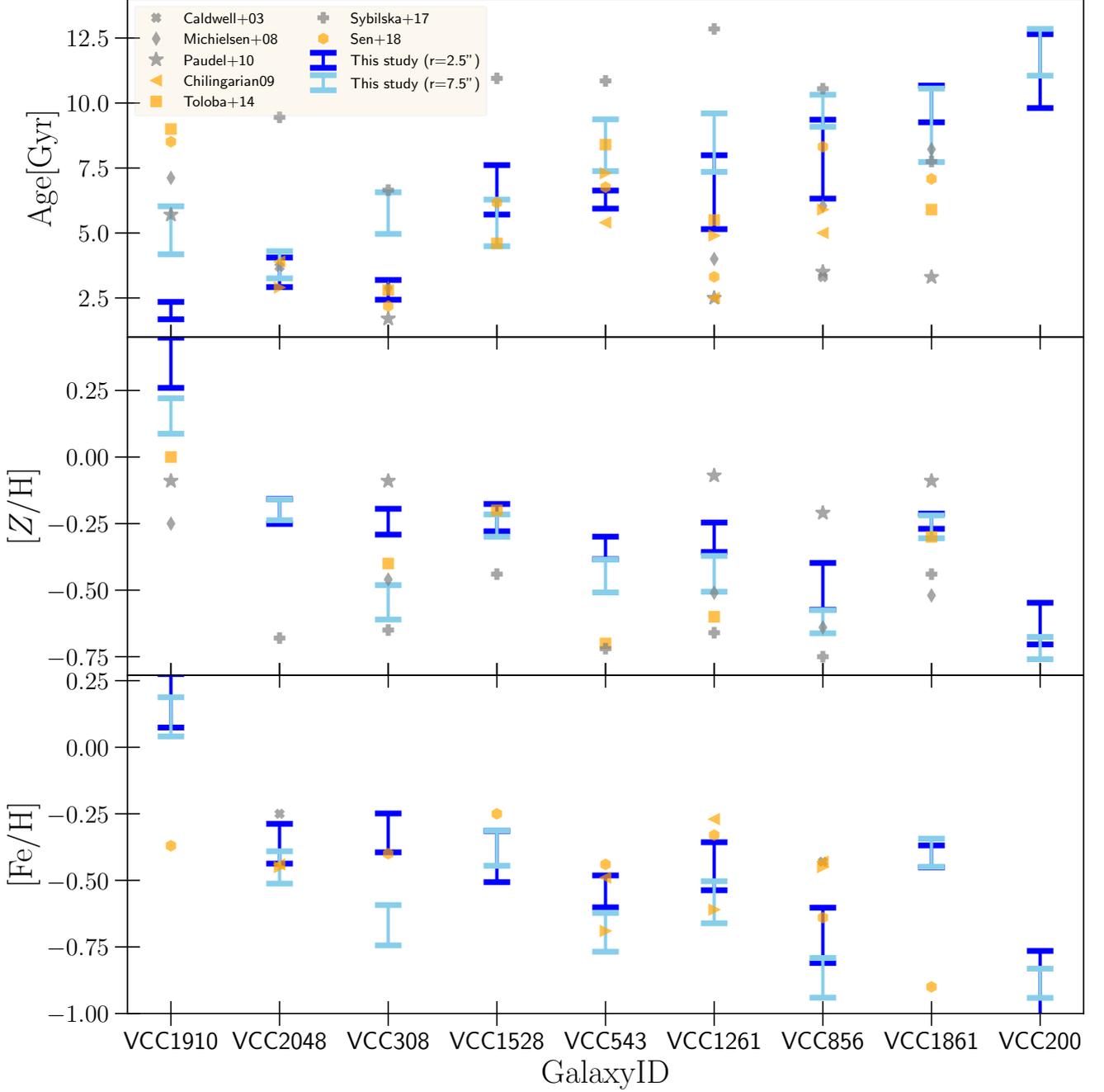


Figure 24. Literature comparison of single stellar population properties. We ordered the galaxies from left to right by the average age we derived in the two apertures. (*Top:* Age in Gyr, *Middle:* Metallicity $[Z/H]$, *Bottom:* Iron abundance $[Fe/H]$ for the studies that trace metallicity in terms of $[Fe/H]$). We use $[Z/H] = [Fe/H] + 0.94 \cdot [\alpha/Fe]$ (Thomas et al. 2003) to convert our metallicity $[Z/H]$ to $[Fe/H]$. The *blue* errorbars show our results for the central spectrum at $r = 2.5''$ (*dark blue*) and the outer one at $r = 7.5''$ (*light blue*). Studies with a spectral resolution of $R \gtrsim 2000$ are highlighted in *orange*, error bars from the literature are not shown to avoid overcrowding (they are generally larger than our errors). Our results for the central spectrum of VCC 1910 are not trustworthy (see Sec. 2.3). For the values of Chilingarian (2009) we plotted the values for the circum-nuclear region and the outer region and indicate the region by the direction of the arrow. We excluded the results of Ryś et al. (2013, 2015) from this plot as they are models composed of two different stellar population models, although we note that they employed the same data as Sybilka et al. (2017). We are not aware of any existing analysis for VCC 200.

such two-burst models are the first step to modeling a possibly more complex star formation history. In this context, we will discuss their results in Sec. 5.3.

Sybilska et al. (2017) re-analyze the data of Ryś et al. (2013) using the same observation and reduction setup but adding more dEs (including VCC 543, VCC 856, VCC 1528). Among other quantities, they derived global dispersions of the *integrated* spectra within $1 r_e$. Similar to the original results of Ryś et al. (2013), the dispersions are higher than the average dispersions we measured, especially so for the lower- σ galaxies of our sample. Unlike Ryś et al. (2013, 2015), they derived population parameters using Lick indices with a *single* population instead of a two burst model, which makes the comparison with our population parameters more meaningful. Compared to many previous studies their results for the integrated spectra within $1 r_e$ yielded very high ages and lower metallicities. In fact, many of the dEs that are also in our sample have no upper constraint for the age as the associated errors reach the edge of their parameter grid (14 Gyr).

The kinematic comparison with the existing literature suggests LOSVDs from lower resolution spectra seem to be artificially broadened especially when the galaxy’s actual dispersion is lower or at the level of the instrumental dispersion. In some cases, low resolution effects may be mitigated by a very high signal-to-noise ratio (Toloba et al. 2011) but this requires that the true noise level in the data is estimated accurately in the first place. Furthermore a too low signal-to-noise of the binned spectra leads to large bin-to-bin scatter which in turn blurs spatial any systematic spatial signatures like the σ -drops present in VCC 1261 and VCC 2048.

We conclude that in the case of the dEs where the dispersion can be as low as 20 km s^{-1} , a resolution $R \gtrsim 8000$ with a $S/N \gtrsim 15$ is required to obtain the width of the LOSVDs in an unbiased manner. For our Virgo galaxies this study is the first to use high-resolution, high S/N , spatially resolved spectra. This allows an unbiased dispersion recovery all the while covering the full 2D kinematic information. It reveals that many of the galaxies have significantly lower dispersions than previous studies suggested. Given the resolution of VIRUS-W we should be able to measure dispersions down to as low as $\sim 15 \text{ km s}^{-1}$, yet the homogeneous dispersion profiles of some galaxies (VCC 200 and VCC 308) could be suggestive that the limit may be reached already at $\sim 20 \text{ km s}^{-1}$. We tested this two-fold. Firstly by correctly recovering a (gaussian) toy LOSVD with $\sigma = 15 \text{ km s}^{-1}$ from a Monte-Carlo mock spectrum of a single stellar population with the same resolution and S/N as the VCC 200 observations. And secondly by adding the *real* spectra of VCC 200 and VCC 308 into larger bins, thus, doubling the S/N per bin. Fitting these spectra did not change the dispersions as σ remained consistent within $\lesssim 0.5 \text{ km s}^{-1}$ with the lower S/N spectra. From these two tests we conclude the dispersions of these two galaxies are robust and the resolution limit is not yet reached.

Concerning stellar population properties, the results seem broadly in agreement with those of the literature as long as we disregard our results for VCC 1910 (see Sec. 2.3). The metallicity values of the dEs are all sub-solar $[Z/H] \in [-0.75, -0.2]$ and span a large range in ages from 3 to 12 Gyr. Both age and metallicity appear to be slightly correlated. Some of this correlation could still be an artifact of the age–metallicity degeneracy, however, the independent measurements of the two spectra at $2.5''$ and $7.5''$ seem broadly consistent for a given galaxy suggesting that they are robust and that the dEs have a radially homogeneous stellar population. Only VCC 308 seems to be significantly younger and metal-rich in its center. This is expected because VCC 308 is the only galaxy in our sample with an extended blue center and as such classified as a dE(bc), while the other dEs do not show color gradients (outside their blue nuclei which are not well resolved; see App. E). We note, however, that many of the literature results have large 1σ error bars and that Paudel et al. (2010), for example, found systematically younger and metal-rich galaxies, whereas Sybilska et al. (2017) find the opposite: systematically older and more metal-poor populations. This may suggest that the age–metallicity degeneracy is not always broken sufficiently. While the population studies with an intermediate spectral resolution (highlighted in yellow in Fig. 24) tend to agree better with our results than those with lower resolution, there is no strong systematic trend, e.g. that low resolution spectra would yield population properties that are biased in a specific direction. We suspect the signal-to-noise of the spectra plays the more important role if one wants to recover unbiased stellar population properties.

C. DO POPULATION MODELS OVERESTIMATE THE TOTAL DYNAMICAL MASS?

For a few of the galaxies in Fig. 17 the IMF parameter Δ_{tot} is slightly negative which suggests that (locally) the total mass predicted dynamically is *lower* than the population models imply. Assuming both mass-to-light ratio estimates are accurate this implies that the IMF of these galaxies is lighter than Kroupa even if we assume they have no dark matter. While this is not per se an issue this could hint at a problem in either the population or dynamical estimates of the galaxies with negative Δ_{tot} . In the following we will discuss possible issues.

Firstly, the negative Δ_{tot} does not necessarily imply that the total, cumulative with radius, dynamical mass is in fact lower than the SSP predictions. The dynamical models we employ have much more flexibility to spatially vary their M_{tot}/L -profiles thanks to the 5 halo parameters (eq. 2) and the log-linear stellar mass-to-light ratios. In contrast, the SSP models are fixed to the light distribution at the integrated spectra around the two probed radii $r = 2.5''$ and $r = 7.5''$. Consequently, the dynamical models are able to change the mass-to-light ratios at all other radii much more easily which could (occasionally) produce negative Δ_{tot} at the two radii evaluated by the inflexible SSP models. To test this we have also dynamically modelled simple mass-follows-light models that have no dark matter component and only a single global mass-to-light ratio sampled in ~ 0.2 steps, and as such these dynamical models are much more comparable in their flexibility to the SSP models. Overall the best mass-follows-light models, shown in Fig. 15, are consistent with the mass-to-light ratios recovered with the more flexible dynamical models that included dark matter and stellar gradients. For most of the galaxies (VCC 308, VCC 1861) that have negative Δ_{tot} , this explanation lifts some of the tension regarding the negative Δ_{tot} . For the remaining galaxy (VCC 200) with $\Delta_{\text{tot}} < 0$, the total mass recovery of the Schwarzschild models could be biased low, or the mass-to-light ratio obtained from the stellar population analysis could be biased high.

None of our application of the dynamical modeling on simulations (e.g. Lipka & Thomas 2021; Neureiter et al. 2023a) have suggested that the *total* dynamical mass could be systematically *underestimated*. If anything it would be easier to bias dynamical masses higher instead of lower. For example, spectra with too low of a resolution would overestimate the observed velocity dispersion (cf. App. B) and, as a consequence, also the inferred dynamical mass. Of course the dynamical models assume the systems are in dynamical equilibrium, which could affect our mass inference if they are not. However, given the old age and photometric structure of VCC 200 (Ferrarese et al. 2006) we have no particularly strong reason to believe this galaxy is more out of equilibrium than the other dEs in the sample. Instead the very old age of VCC 200 (12 Gyr) could be slightly overestimated, which could bias its SSP mass-to-light ratio Υ_{Kroupa} higher, resulting in a negative Δ_{tot} . Unfortunately, VCC 200 is also the only galaxy in our sample for which no reference SSP results exist in the literature so we cannot confirm its age. However, considering its low $H\beta$ index ($\lesssim 2.0$) the galaxy is very likely older than 9 Gyr. In conclusion, we suspect the age of VCC 200 is either slightly overestimated by 1 or 2 Gyr and consistent with Kroupa IMF or, alternatively, the age is correct and the galaxy has a IMF slightly lighter than Kroupa.

D. IMF–METALLICITY COUPLING

Fig. 25 shows the relation of IMF parameter with the metallicity. Apart from VCC 1861 the IMF parameter is positively correlated with $[Z/H]$. If the Δ_* we derived are robust, it could be that either age or metallicity (or both together) are the physical reason behind the variety seen in the IMF of dEs. For ordinary ETGs strong, positive correlation of the IMF parameter and the metallicity have been noticed before (e.g. Martín-Navarro et al. 2015; Parikh et al. 2018; Li et al. 2023).

E. PHENOMENOLOGICAL DISCUSSION OF EACH GALAXY - COLORS, SUBSTRUCTURE, ENVIRONMENT, KINEMATIC SIGNATURES

In this section, we discuss each dE holistically based solely on the photometric and kinematic data we derived and what we found in the existing literature. A detailed description of the *HST* photometry, isophotes and colors can be found in Ferrarese et al. (2006) for all galaxies in our sample, except for VCC 308. In the following, unless stated otherwise, all colors are stated in $g - z$ bands.

VCC 200: Classified as a dE2(N), the ellipticity structure is actually fairly round for large parts of the galaxy, only experiencing a double peak up to $\epsilon \sim 0.2$ at $0.5''$ and $5''$ (Ferrarese et al. 2006). This signature may stem from a ring structure inhabiting an otherwise rounder galaxy. Boxiness/diskiness parameters scatter but are otherwise consistent with 0. The $g - z$ color is constant with radius at $1.2 - 1.3$ mag. Only in the center, it has a slightly bluer (1.1 mag) and bright nucleus (cf. Hamraz et al. 2019). We find the galaxy exhibits intermediate rotation (≤ 10 km s $^{-1}$) and a constant, low dispersion (~ 25 km s $^{-1}$) within the VIRUS-W FoV. While one may discern a hint of $v-h_3$ anticorrelation in the maps, even higher Gauss–Hermite moments scatter significantly (but stay within ± 0.1). If the heliocentric velocity and distance to us and M87 are to be believed, VCC 200 is at the backside of the cluster far away from the center and is moving towards it. This stands in contrast to its very old SSP age (Sec. 2.3). Within our sample, VCC 200 seems to be an outlier in regard to its stellar population, with it being very old (> 10 Gyr) and metal-poor compared to the rest of our results. Unfortunately, it is also the only galaxy for which we are not aware of any existing kinematic or stellar population studies.

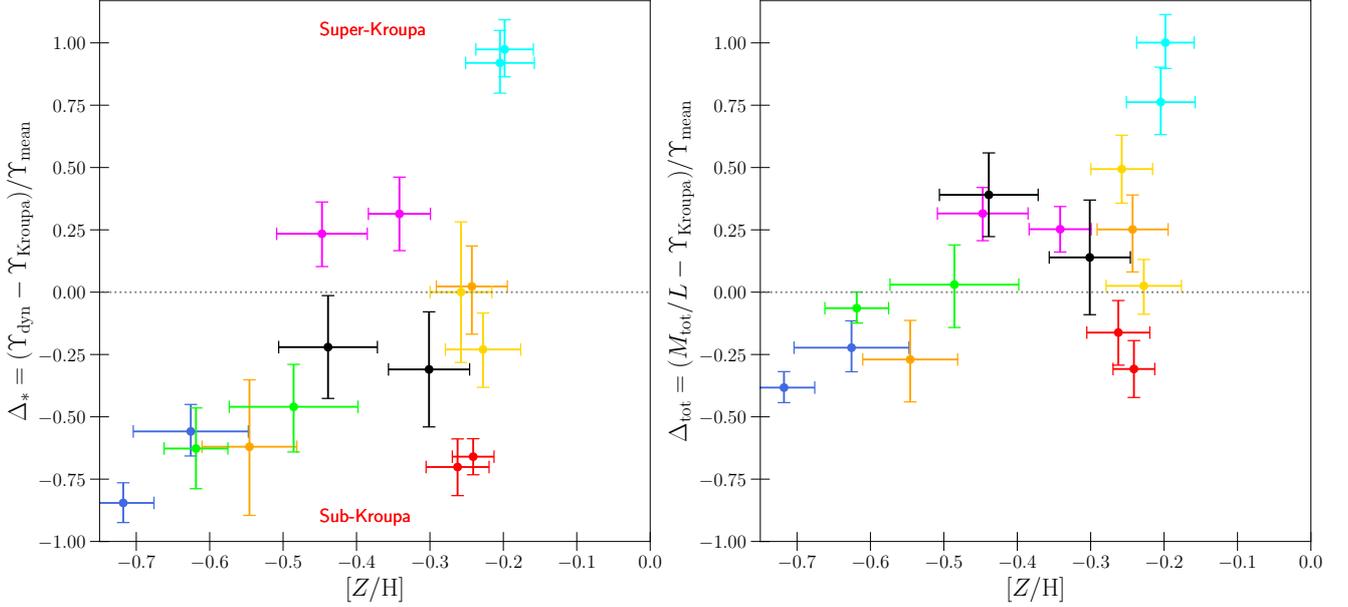


Figure 25. Analogous to Fig. 17 but for the metallicity instead of the age. VCC 1910 is not in the plot range due to its (likely wrong) super-solar metallicity. Due to the anti-correlation of age and metallicity (Fig. 4) of our dEs the positive correlation of the IMF parameter and the metallicity is a corollary of the IMF and age relation.

VCC 308: VCC 308 is the only galaxy in our sample classified as a dE(bc) (cf. Lisker et al. 2007), meaning it is significantly bluer in the center, showing an extended radial color gradient becoming redder further outside (i.e. not just a distinct blue nucleus). However, in the case of VCC 308 this gradient is relatively small in terms of radial extent and magnitude (cf. Lisker et al. 2006a). One could debate whether it is really that distinct to the rest of our sample galaxies. Lisker et al. (2006b) also find the photometry shows weak signs of spiral arms that are seen face-on. The kinematics of VCC 308 we measured is similar to that of VCC 200. It has intermediate rotation while having the lowest dispersion values ($20 - 25 \text{ km s}^{-1}$) in our sample (still significantly higher than the resolution of VIRUS-W). Higher Gauss-Hermite moments are noisy, but we identify a hint of a $v - h_3$ anti-correlation and a positive h_4 within the vast majority of the Voronoi bins. VCC 308 is far away from the center in a very low projected density environment (Sybilka et al. 2017).

VCC 543: VCC 543 appears to be fairly elongated, being classified as a dE5. While the designation by Lisker et al. (2007) suggests no nucleus, we and Hamraz et al. (2019); Ferrarese et al. (2006) find a detectable, round, and slightly bluer nucleus in the galaxies center. Other than that, the galaxy is quite regular, showing no preference for either diskiness nor boxiness or any detectable substructures. The galaxy shows a strong, linearly rising, velocity signal that is anti-correlated with h_3 . Within $10''$, h_4 appears to experience a radial drop off, however, further outside the signal becomes quite noisy, showing no clear trend. The dispersion increases with radius, showing no signs of plateauing within the FoV. The galaxy is in the foreground of the cluster, moving slightly away from it, possibly because it has previously passed the cluster center.

VCC 856: At first glance the galaxy appears to be a typical, fairly round dE ($\epsilon \sim 0.1$) with a $g-z$ -color of ~ 1.2 mag and a bluer (~ 1.0 mag) nucleus that dominates its center. However, as first noted by Jerjen et al. (2000), one can see a faint signature of face-on spiral arms in the galaxy and as such the classification of VCC 856 as a dE is debatable. Perhaps we see the galaxy during its transformation from a dwarf spiral into a dE. Despite the *presumably* face-on disk, the galaxy displays clear signs of rotation around its axis, suggesting at least some degree of inclination. The dispersion is generally low, with a *decreasing* dispersion. It stands out significantly from the rest of the dE sample by exhibiting a strongly *rising* h_4 profile with comparatively larger error bars. The scatter in h_3 is large, possibly due to the spiral arms.

VCC 1261: VCC 1261 is the brightest galaxy in our sample. Despite its significant flattening ($\epsilon \sim 0.4$) the galaxy's isophotes are regular ellipticals ($a_4 \sim 0$) showing no signs of an embedded disk. In its center it hosts a rounder, bluer (1.0 mag compared to 1.2 mag) nucleus. Within $\sim 10''$ the kinematics show little rotation but a strongly rising

dispersion reaching $\sigma = 50 \text{ km s}^{-1}$. Accompanied by this dispersion rise is a significant drop in h_4 down to 0 at $10''$, suggesting a strong σ - h_4 -correlation. In the outermost radial bins the galaxy behaves quite differently with the dispersion and h_4 plateauing, while the rotation abruptly becomes more significant. The orbit of VCC 1261 within Virgo may be odd, as it seems to be far behind the cluster yet is still moving away from Virgo relatively rapidly.

The fact that VCC 1261 appears to be a very flattened galaxy, yet having essentially no rotation in our and most other kinematic studies (Geha et al. 2003; van Zee et al. 2004; Chilingarian 2009; Ryś et al. 2013; Toloba et al. 2014, 2015; Sybilska et al. 2017; Şen et al. 2018), seems to be in conflict with the hypothesis that that dEs are remnants of more disk, gas-rich, late-type dwarfs. However, in our and other studies we observe a rise in velocity beyond $12''$ which could suggest that we simply do not probe the galaxy at large enough radii to see a more pronounced rotation signal. The same could apply to the other non-rotator in our sample, VCC 1528. Both non-rotators are, perhaps not coincidentally, the two dEs nearest to the cluster center (Tab. 1). Initially their mass and extent could have been larger than the other dEs in our sample, but their increased likelihood of interactions via harassment and ram-pressure stripping could have made the two dEs significantly fainter at larger radii, effectively ‘shrinking’ them to the regime of our sample. This appears fairly plausible because we will find that these two galaxies have a significantly more massive dark matter halo than the other dEs (cf. VW–II). This scenario is also suggested by Beasley et al. (2009) who studied the kinematics of VCC 1261’s globular clusters (GC), which, unlike studies based on the galaxy’s integrated light, allows constraints on v and σ much farther out at several effective radii. However, this also comes with fairly large uncertainties and the assumption that the galaxy’s stars and GCs are closely kinematically associated. The GC motions they found suggest significant rotation ($v/\sigma > 1$) at larger radii. Future IFU studies with deeper kinematics may resolve whether the velocity of the stars is indeed rising further beyond $1 r_e$ for the two galaxies.

It could also be that VCC 1261 is an outlier. The E7/S0 galaxy NGC 4550 for example is known to host two co-spatial, *counter-rotating* disks of equal mass which results in a net streaming motion that is extremely low, yet the systems is very flat (Rubin et al. 1992; Rix et al. 1992; Emsellem et al. 2007). A sign that such a counter-rotating disk could be embedded within the central $\sim 10''$ where $v = 0$ is the congruent decrease in σ and increase in h_4 towards the centre (Fig. 3). However, from its orbit structure obtained from the dynamical modeling (Fig. 13 and Fig. 14) we do not find strong evidence in favor of this scenario.

VCC 1528: Despite being classified as non-nucleated dE in Lisker et al. (2007), we and Ferrarese et al. (2006) find the galaxy to host a resolved, blue nucleus. Apart from this nucleus the galaxy is fairly red ($\sim 1.35 \text{ mag}$) near its center, progressively becoming bluer ($\sim 1.2 \text{ mag}$) at larger radii. At large radii ($> 10''$) the galaxy appears fairly round ($\epsilon \sim 0.1$). However, it is significantly more elliptical in the center with $\epsilon \sim 0.25$. The shapes of the ellipses are very regular. Within its small FoV, VCC 1528 is a non-rotator with a virtually flat (or very slightly rising) dispersion profile. Higher Gauss–Hermite moments scatter, especially near the edge of the FoV, but are overall consistent with (and close to) zero. Since the galaxy becomes flatter and bluer outside the FoV, a change in kinematics may be expected. VCC 1528 is the dE closest to Virgo’s 3D center.

VCC 1861: Classified as dE0, the galaxy is very round and, as such, has no well constrained position angle. The color of the galaxy ($\sim 1.3 \text{ mag}$) and the bluer, very bright nucleus are typical for our sample. The isophote ellipses are regular and we find no substructures in the galaxy. VCC 1861 shows intermediate rotation, and a slightly rising dispersion profile within $\sim 10''$, plateauing at larger radii. The LOSVDs are roughly symmetric with only a tentative sign of a v - h_3 anti-correlation in the kinematic map. Similar to VCC 1261 and VCC 2048, the galaxy has strongly peaked LOSVDs in the center ($h_4 \sim 0.1$) which steadily become more Gaussian ($h_4 \sim 0$) with increasing radius. VCC 1861 may be associated with a locally denser sub-clump of galaxies formed around the large elliptical galaxy M60.

VCC 1910: VCC 1910 is the reddest galaxy in our sample with $g - z \approx 1.4 \text{ mag}$, but has a typical blue nucleus with $g - z \approx 1.0 \text{ mag}$. Akin to VCC 200, the ellipticity has two distinct and extended peaks at $0.8''$ and $6''$ reaching $\epsilon \sim 0.2$. VCC 1910 is an intermediate rotator which, together with VCC 856, stands out in that the dispersion peaks in the center and then steadily drops of with radius. In the maps one can again see signs of v - h_3 anti-correlation, but overall higher moments scatter significantly and are consistent with zero. VCC 1910 might also be associated with the M60 sub-clump, but its net velocity suggests it moves relative quickly towards us.

VCC 2048: Together with VCC 543 and VCC 1261 the galaxy is on the bluer end of our sample with a color of 1.2 mag and a blue nucleus at 1.0 mag . And, again together with VCC 543 and VCC 1261, it is significantly elongated ($\epsilon \sim 0.5$ to 0.6). It also hosts a large disk that causes the isophotes to be disk-like over a large radial range. For a more in-depth study of this galaxy’s photometry and a bulge-disk decomposition, we refer the reader to Kormendy & Bender (2012). VCC 2048 shows very strong rotation around its minor axis and, very similar to the other flattened galaxy

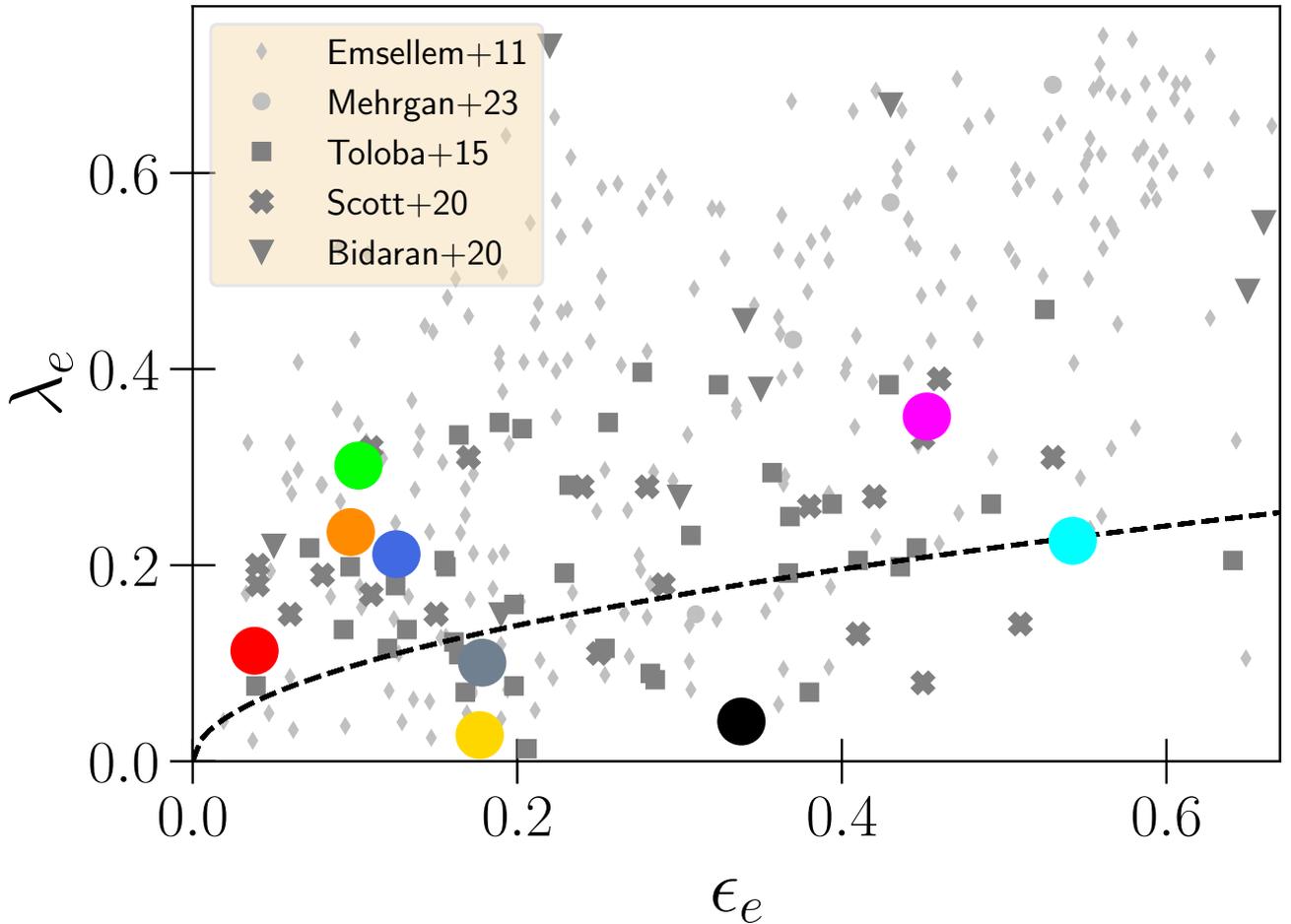


Figure 26. Alternative version of Fig. 5 analysed within 1.0 instead of 0.5 stellar effective radii. For some galaxies in the VIRUS-W sample and in the literature this requires an extrapolation because the kinematic maps do not extend far enough. The typical distribution of dEs does not change much with the increase in aperture apart from two of the infalling dEs of Bidaran et al. (2020) becoming outliers.

VCC 1261, has a strongly rising dispersion profile that plateaus at $\sim 8''$ without signs of a drop-off within the FoV. The velocity and skewness h_3 are clearly anti-correlated, and h_4 has a strong peak in the center which starts to drop off with increasing radius, even becoming slightly negative in the outermost bins. Similar to VCC 308, VCC 2048 is located at a very low projected cluster density (Sybiliska et al. 2017).

F. ANGULAR MOMENTUM VERSUS STELLAR MASS AND λ_E

Fig. 26 and Fig. 27 are alternative versions to Fig. 5 and Fig. 6. The former shows the angular momentum parameter within $1 r_e$ and the latter the specific angular momentum j against stellar mass instead of B -band magnitude. In some of the studies shown in Fig. 27 the stellar masses were not stated, and we proceeded as follows. For the Toloba et al. (2015) dEs we use the stellar masses obtained from Tortora et al. (2019). For the Martínez-García et al. (2021) dSphs we adopted stellar masses from Hayashi et al. (2020). For the Local Group dEs (Geha et al. 2006, 2010) we used values from Mateo (1998). For the ‘ordinary’ ETGs of Bender & Nieto (1990) we estimate a mass from the B -band magnitudes and an assumed mass-to-light ratio of 5; for their dwarfs we again used the values from Mateo (1998) and Hayashi et al. (2020). For the dEs of Geha et al. (2003) we convert from magnitudes and assume the median mass-to-light ratio that was given for a sub-sample of these dEs (see Geha et al. 2002). To estimate the stellar masses of Emsellem et al. (2011) we use eq. 28 of Cappellari et al. (2013a) and the values stated therein

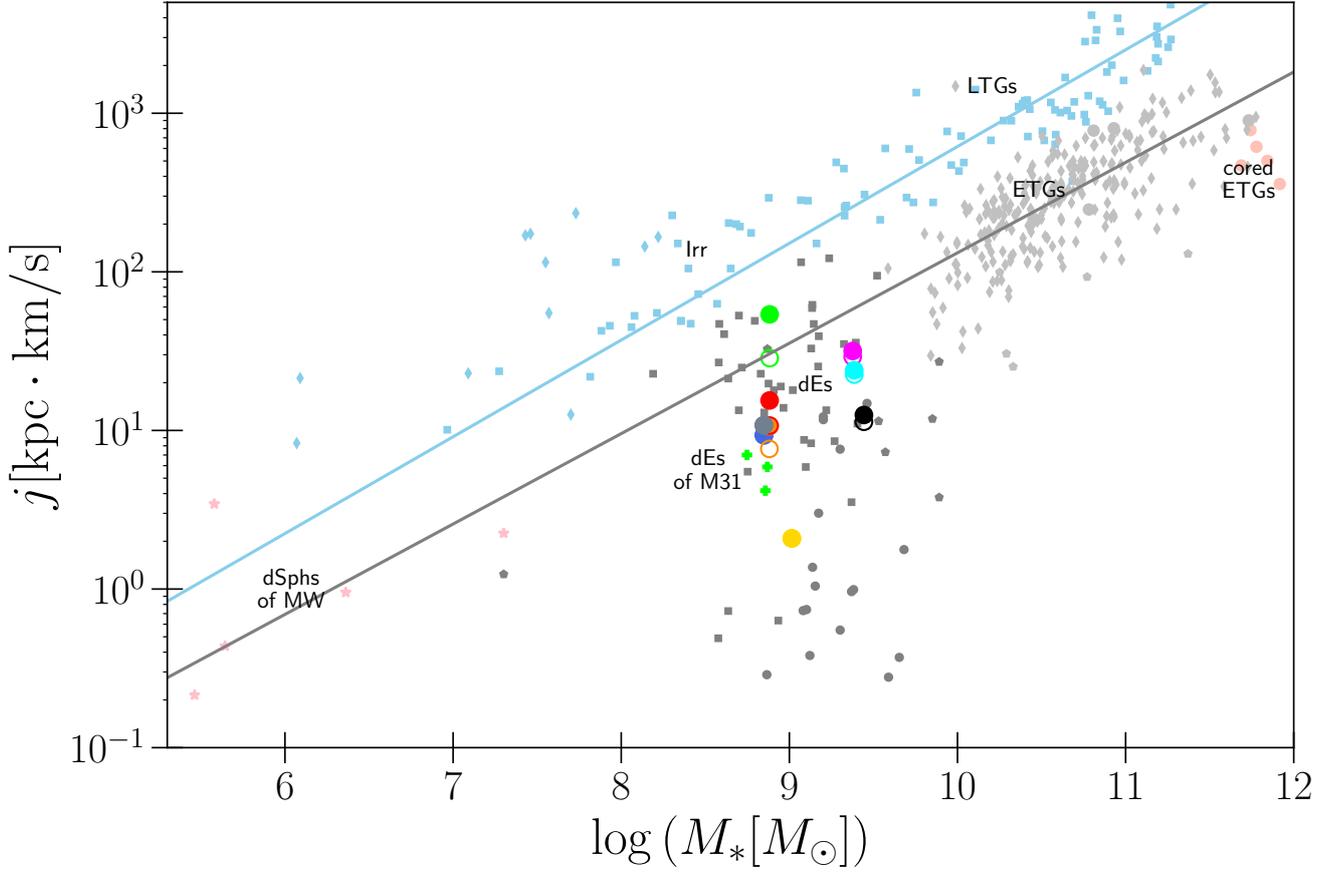


Figure 27. As Fig. 6 but versus the total stellar mass M_* . *Gray diagonal:* Standard $j - M_*$ scaling relation from Pulsoni et al. (2023) for fast rotators, i.e. this is *not* a fit to any of the galaxies shown here. *Blue diagonal:* Standard scaling relation for massive LTGs (Di Teodoro et al. 2023).

REFERENCES

- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, *ApJS*, 203, 21, doi: [10.1088/0067-0049/203/2/21](https://doi.org/10.1088/0067-0049/203/2/21)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Audet, C., & Dennis, J. E. 2006, *SIAM Journal on Optimization*, 17, 188, doi: [10.1137/040603371](https://doi.org/10.1137/040603371)
- Barazza, F. D., Binggeli, B., & Jerjen, H. 2002, *A&A*, 391, 823, doi: [10.1051/0004-6361:20020875](https://doi.org/10.1051/0004-6361:20020875)
- Barnes, J. E., & Hernquist, L. 1992, *Nature*, 360, 715, doi: [10.1038/360715a0](https://doi.org/10.1038/360715a0)
- Bastian, N., Covey, K. R., & Meyer, M. R. 2010, *ARA&A*, 48, 339, doi: [10.1146/annurev-astro-082708-101642](https://doi.org/10.1146/annurev-astro-082708-101642)
- Beasley, M. A., Cenarro, A. J., Strader, J., & Brodie, J. P. 2009, *AJ*, 137, 5146, doi: [10.1088/0004-6256/137/6/5146](https://doi.org/10.1088/0004-6256/137/6/5146)
- Bender, R. 1988, *A&A*, 193, L7
- Bender, R., Burstein, D., & Faber, S. M. 1992, *ApJ*, 399, 462, doi: [10.1086/171940](https://doi.org/10.1086/171940)
- Bender, R., & Nieto, J. L. 1990, *A&A*, 239, 97
- Bender, R., Paquet, A., & Nieto, J. L. 1991, *A&A*, 246, 349
- Bender, R., Saglia, R. P., & Gerhard, O. E. 1994, *MNRAS*, 269, 785, doi: [10.1093/mnras/269.3.785](https://doi.org/10.1093/mnras/269.3.785)
- Bender, R., Surma, P., Doebereiner, S., Moellenhoff, C., & Madejsky, R. 1989, *A&A*, 217, 35
- Bidaran, B., Pasquali, A., Lisker, T., et al. 2020, *MNRAS*, 497, 1904, doi: [10.1093/mnras/staa2097](https://doi.org/10.1093/mnras/staa2097)
- Bidaran, B., La Barbera, F., Pasquali, A., et al. 2022, *MNRAS*, 515, 4622, doi: [10.1093/mnras/stac2005](https://doi.org/10.1093/mnras/stac2005)
- . 2023, *MNRAS*, 525, 4329, doi: [10.1093/mnras/stad2546](https://doi.org/10.1093/mnras/stad2546)
- Binggeli, B., Sandage, A., & Tammann, G. A. 1985, *AJ*, 90, 1681, doi: [10.1086/113874](https://doi.org/10.1086/113874)
- Binggeli, B., Tammann, G. A., & Sandage, A. 1987, *AJ*, 94, 251, doi: [10.1086/114467](https://doi.org/10.1086/114467)
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* (Princeton University Press)
- Blakeslee, J. P., Jordán, A., Mei, S., et al. 2009, *ApJ*, 694, 556, doi: [10.1088/0004-637X/694/1/556](https://doi.org/10.1088/0004-637X/694/1/556)

- Boch, T., & Fernique, P. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 485, *Astronomical Data Analysis Software and Systems XXIII*, ed. N. Manset & P. Forshay, 277
- Boselli, A., Fossati, M., & Sun, M. 2022, *A&A Rv*, 30, 3, doi: [10.1007/s00159-022-00140-3](https://doi.org/10.1007/s00159-022-00140-3)
- Brewer, B. J., Dutton, A. A., Treu, T., et al. 2012, *MNRAS*, 422, 3574, doi: [10.1111/j.1365-2966.2012.20870.x](https://doi.org/10.1111/j.1365-2966.2012.20870.x)
- Burnham, K., & Anderson, D. 2002, *Model selection and multimodel inference: a practical information-theoretic approach* (Springer Verlag)
- Butler, K. M., Obreschkow, D., & Oh, S.-H. 2017, *ApJL*, 834, L4, doi: [10.3847/2041-8213/834/1/L4](https://doi.org/10.3847/2041-8213/834/1/L4)
- Caldwell, N., Rose, J. A., & Concannon, K. D. 2003, *AJ*, 125, 2891, doi: [10.1086/375308](https://doi.org/10.1086/375308)
- Cappellari, M. 2016, *ARA&A*, 54, 597, doi: [10.1146/annurev-astro-082214-122432](https://doi.org/10.1146/annurev-astro-082214-122432)
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, 342, 345, doi: [10.1046/j.1365-8711.2003.06541.x](https://doi.org/10.1046/j.1365-8711.2003.06541.x)
- Cappellari, M., Emsellem, E., Bacon, R., et al. 2007, *MNRAS*, 379, 418, doi: [10.1111/j.1365-2966.2007.11963.x](https://doi.org/10.1111/j.1365-2966.2007.11963.x)
- Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, 413, 813, doi: [10.1111/j.1365-2966.2010.18174.x](https://doi.org/10.1111/j.1365-2966.2010.18174.x)
- Cappellari, M., Scott, N., Alatalo, K., et al. 2013a, *MNRAS*, 432, 1709, doi: [10.1093/mnras/stt562](https://doi.org/10.1093/mnras/stt562)
- Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2013b, *MNRAS*, 432, 1862, doi: [10.1093/mnras/stt644](https://doi.org/10.1093/mnras/stt644)
- Chabrier, G. 2003, *PASP*, 115, 763, doi: [10.1086/376392](https://doi.org/10.1086/376392)
- Chilingarian, I. V. 2009, *Monthly Notices of the Royal Astronomical Society*, 394, 1229, doi: [10.1111/j.1365-2966.2009.14450.x](https://doi.org/10.1111/j.1365-2966.2009.14450.x)
- Chon, S., Hosokawa, T., Omukai, K., & Schneider, R. 2024, *MNRAS*, 530, 2453, doi: [10.1093/mnras/stae1027](https://doi.org/10.1093/mnras/stae1027)
- Chon, S., Ono, H., Omukai, K., & Schneider, R. 2022, *MNRAS*, 514, 4639, doi: [10.1093/mnras/stac1549](https://doi.org/10.1093/mnras/stac1549)
- Chowdhury, A., & Chengalur, J. N. 2017, *MNRAS*, 467, 3856, doi: [10.1093/mnras/stx355](https://doi.org/10.1093/mnras/stx355)
- Sen, S., Peletier, R. F., Boselli, A., et al. 2018, *MNRAS*, 475, 3453, doi: [10.1093/mnras/stx3254](https://doi.org/10.1093/mnras/stx3254)
- De Leo, M., Read, J. I., Noel, N. E. D., et al. 2023, *arXiv e-prints*, arXiv:2303.08838, doi: [10.48550/arXiv.2303.08838](https://doi.org/10.48550/arXiv.2303.08838)
- de los Reyes, M. A. C., Kirby, E. N., Zhuang, Z., et al. 2023, *ApJ*, 951, 52, doi: [10.3847/1538-4357/acd189](https://doi.org/10.3847/1538-4357/acd189)
- de Nicola, S., Saglia, R. P., Thomas, J., Dehnen, W., & Bender, R. 2020, *MNRAS*, 496, 3076, doi: [10.1093/mnras/staa1703](https://doi.org/10.1093/mnras/staa1703)
- De Rijcke, S., Prugniel, P., Simien, F., & Dejonghe, H. 2006, *MNRAS*, 369, 1321, doi: [10.1111/j.1365-2966.2006.10377.x](https://doi.org/10.1111/j.1365-2966.2006.10377.x)
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, 368, 2, doi: [10.1111/j.1365-2966.2006.10145.x](https://doi.org/10.1111/j.1365-2966.2006.10145.x)
- Di Teodoro, E. M., Posti, L., Fall, S. M., et al. 2023, *MNRAS*, 518, 6340, doi: [10.1093/mnras/stac3424](https://doi.org/10.1093/mnras/stac3424)
- Ding, Y., Zhu, L., van de Ven, G., et al. 2023, *A&A*, 672, A84, doi: [10.1051/0004-6361/202244558](https://doi.org/10.1051/0004-6361/202244558)
- Dressler, A. 1980, *ApJ*, 236, 351, doi: [10.1086/157753](https://doi.org/10.1086/157753)
- Efstathiou, G., & Jones, B. J. T. 1979, *MNRAS*, 186, 133, doi: [10.1093/mnras/186.2.133](https://doi.org/10.1093/mnras/186.2.133)
- Eftekhari, F. S., Peletier, R. F., Scott, N., et al. 2022, *MNRAS*, 517, 4714, doi: [10.1093/mnras/stac2606](https://doi.org/10.1093/mnras/stac2606)
- Emsellem, E., Cappellari, M., Krajnović, D., et al. 2007, *MNRAS*, 379, 401, doi: [10.1111/j.1365-2966.2007.11752.x](https://doi.org/10.1111/j.1365-2966.2007.11752.x)
- . 2011, *MNRAS*, 414, 888, doi: [10.1111/j.1365-2966.2011.18496.x](https://doi.org/10.1111/j.1365-2966.2011.18496.x)
- Ene, I., Ma, C.-P., Veale, M., et al. 2018, *MNRAS*, 479, 2810, doi: [10.1093/mnras/sty1649](https://doi.org/10.1093/mnras/sty1649)
- Fabricius, M. H., Barnes, S., Bender, R., et al. 2008, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7014, *Ground-based and Airborne Instrumentation for Astronomy II*, ed. I. S. McLean & M. M. Casali, 701473, doi: [10.1117/12.787204](https://doi.org/10.1117/12.787204)
- Fabricius, M. H., Grupp, F., Bender, R., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, *Ground-based and Airborne Instrumentation for Astronomy IV*, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84465K, doi: [10.1117/12.925177](https://doi.org/10.1117/12.925177)
- Falcón-Barroso, J., & Martig, M. 2021, *A&A*, 646, A31, doi: [10.1051/0004-6361/202039624](https://doi.org/10.1051/0004-6361/202039624)
- Falcón-Barroso, J., van de Ven, G., Lyubenova, M., et al. 2019, *A&A*, 632, A59, doi: [10.1051/0004-6361/201936413](https://doi.org/10.1051/0004-6361/201936413)
- Fall, S. M. 1983, in *Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula, Vol. 100, 391–398
- Ferrarese, L., Côté, P., Jordán, A., et al. 2006, *ApJS*, 164, 334, doi: [10.1086/501350](https://doi.org/10.1086/501350)
- Ferrarese, L., Côté, P., Cuillandre, J.-C., et al. 2012, *ApJS*, 200, 4, doi: [10.1088/0067-0049/200/1/4](https://doi.org/10.1088/0067-0049/200/1/4)
- Ferreras, I., Barbera, F. L., Rosa, I. G. d. l., et al. 2012, *Monthly Notices of the Royal Astronomical Society: Letters*, 429, L15, doi: [10.1093/mnrasl/sls014](https://doi.org/10.1093/mnrasl/sls014)
- Frigo, M., Naab, T., Rantala, A., et al. 2021, *MNRAS*, 508, 4610, doi: [10.1093/mnras/stab2754](https://doi.org/10.1093/mnras/stab2754)
- Fujita, Y. 2004, *PASJ*, 56, 29, doi: [10.1093/pasj/56.1.29](https://doi.org/10.1093/pasj/56.1.29)
- Geha, M., Blanton, M. R., Yan, R., & Tinker, J. L. 2012, *ApJ*, 757, 85, doi: [10.1088/0004-637X/757/1/85](https://doi.org/10.1088/0004-637X/757/1/85)
- Geha, M., Guhathakurta, P., Rich, R. M., & Cooper, M. C. 2006, *AJ*, 131, 332, doi: [10.1086/498686](https://doi.org/10.1086/498686)
- Geha, M., Guhathakurta, P., & van der Marel, R. P. 2002, *AJ*, 124, 3073, doi: [10.1086/344764](https://doi.org/10.1086/344764)

- . 2003, *AJ*, 126, 1794, doi: [10.1086/377624](https://doi.org/10.1086/377624)
- Geha, M., van der Marel, R. P., Guhathakurta, P., et al. 2010, *ApJ*, 711, 361, doi: [10.1088/0004-637X/711/1/361](https://doi.org/10.1088/0004-637X/711/1/361)
- Gerhard, O., Kronawitter, A., Saglia, R. P., & Bender, R. 2001, *AJ*, 121, 1936, doi: [10.1086/319940](https://doi.org/10.1086/319940)
- Gerhard, O. E., & Binney, J. J. 1996, *MNRAS*, 279, 993, doi: [10.1093/mnras/279.3.993](https://doi.org/10.1093/mnras/279.3.993)
- Gössl, C. A., Drory, N., Relke, H., et al. 2006, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 6270, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ed. D. R. Silva & R. E. Doxsey, 627021, doi: [10.1117/12.671264](https://doi.org/10.1117/12.671264)
- Gössl, C. A., & Riffeser, A. 2002, *A&A*, 381, 1095, doi: [10.1051/0004-6361:20011522](https://doi.org/10.1051/0004-6361:20011522)
- Graham, M. T., Cappellari, M., Li, H., et al. 2018, *MNRAS*, 477, 4711, doi: [10.1093/mnras/sty504](https://doi.org/10.1093/mnras/sty504)
- Gunn, J. E., & Gott, J. Richard, I. 1972, *ApJ*, 176, 1, doi: [10.1086/151605](https://doi.org/10.1086/151605)
- Guo, K., Cortese, L., Obreschkow, D., et al. 2020, *MNRAS*, 491, 773, doi: [10.1093/mnras/stz3042](https://doi.org/10.1093/mnras/stz3042)
- Hamraz, E., Peletier, R. F., Khosroshahi, H. G., et al. 2019, *A&A*, 625, A94, doi: [10.1051/0004-6361/201935076](https://doi.org/10.1051/0004-6361/201935076)
- Hayashi, K., Chiba, M., & Ishiyama, T. 2020, *ApJ*, 904, 45, doi: [10.3847/1538-4357/abbe0a](https://doi.org/10.3847/1538-4357/abbe0a)
- Hill, G. J., Gebhardt, K., Komatsu, E., & MacQueen, P. J. 2004, in *American Institute of Physics Conference Series*, Vol. 743, *The New Cosmology: Conference on Strings and Cosmology*, ed. R. E. Allen, D. V. Nanopoulos, & C. N. Pope, 224–233, doi: [10.1063/1.1848329](https://doi.org/10.1063/1.1848329)
- Hill, G. J., Lee, H., MacQueen, P. J., et al. 2021, *AJ*, 162, 298, doi: [10.3847/1538-3881/ac2c02](https://doi.org/10.3847/1538-3881/ac2c02)
- Houghton, R. C. W., Magorrian, J., Sarzi, M., et al. 2006, *MNRAS*, 367, 2, doi: [10.1111/j.1365-2966.2005.09713.x](https://doi.org/10.1111/j.1365-2966.2005.09713.x)
- Janz, J., Penny, S. J., Graham, A. W., Forbes, D. A., & Davies, R. L. 2017, *MNRAS*, 468, 2850, doi: [10.1093/mnras/stx634](https://doi.org/10.1093/mnras/stx634)
- Jardel, J. R., & Gebhardt, K. 2012, *ApJ*, 746, 89, doi: [10.1088/0004-637X/746/1/89](https://doi.org/10.1088/0004-637X/746/1/89)
- Jardel, J. R., Gebhardt, K., Fabricius, M. H., Drory, N., & Williams, M. J. 2013, *ApJ*, 763, 91, doi: [10.1088/0004-637X/763/2/91](https://doi.org/10.1088/0004-637X/763/2/91)
- Jerjen, H., Kalnajs, A., & Binggeli, B. 2000, *A&A*, 358, 845. <https://arxiv.org/abs/astro-ph/0004248>
- Jin, Y., Zhu, L., Long, R. J., et al. 2020, *MNRAS*, 491, 1690, doi: [10.1093/mnras/stz3072](https://doi.org/10.1093/mnras/stz3072)
- . 2019, *MNRAS*, 486, 4753, doi: [10.1093/mnras/stz1170](https://doi.org/10.1093/mnras/stz1170)
- Klimentowski, J., Lokas, E. L., Kazantzidis, S., Mayer, L., & Mamon, G. A. 2009, *MNRAS*, 397, 2015, doi: [10.1111/j.1365-2966.2009.15046.x](https://doi.org/10.1111/j.1365-2966.2009.15046.x)
- Kormendy, J. 1985, *ApJ*, 295, 73, doi: [10.1086/163350](https://doi.org/10.1086/163350)
- Kormendy, J. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 182, *Galaxy Dynamics - A Rutgers Symposium*, ed. D. R. Merritt, M. Valluri, & J. A. Sellwood, 124
- Kormendy, J., & Bender, R. 1996, *ApJL*, 464, L119, doi: [10.1086/310095](https://doi.org/10.1086/310095)
- . 2009, *ApJL*, 691, L142, doi: [10.1088/0004-637X/691/2/L142](https://doi.org/10.1088/0004-637X/691/2/L142)
- . 2012, *ApJS*, 198, 2, doi: [10.1088/0067-0049/198/1/2](https://doi.org/10.1088/0067-0049/198/1/2)
- . 2013, *ApJL*, 769, L5, doi: [10.1088/2041-8205/769/1/L5](https://doi.org/10.1088/2041-8205/769/1/L5)
- Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, *ApJS*, 182, 216, doi: [10.1088/0067-0049/182/1/216](https://doi.org/10.1088/0067-0049/182/1/216)
- Kormendy, J., & Freeman, K. C. 2016, *ApJ*, 817, 84, doi: [10.3847/0004-637X/817/2/84](https://doi.org/10.3847/0004-637X/817/2/84)
- Kowalczyk, K., & Lokas, E. L. 2022, *A&A*, 659, A119, doi: [10.1051/0004-6361/202142212](https://doi.org/10.1051/0004-6361/202142212)
- Kroupa, P. 2001, *MNRAS*, 322, 231, doi: [10.1046/j.1365-8711.2001.04022.x](https://doi.org/10.1046/j.1365-8711.2001.04022.x)
- . 2002, *Science*, 295, 82, doi: [10.1126/science.1067524](https://doi.org/10.1126/science.1067524)
- Kurapati, S., Chengalur, J. N., Pustilnik, S., & Kamphuis, P. 2018, *MNRAS*, 479, 228, doi: [10.1093/mnras/sty1397](https://doi.org/10.1093/mnras/sty1397)
- La Barbera, F., Ferreras, I., Vazdekis, A., et al. 2013, *MNRAS*, 433, 3017, doi: [10.1093/mnras/stt943](https://doi.org/10.1093/mnras/stt943)
- Larson, R. B. 2005, *MNRAS*, 359, 211, doi: [10.1111/j.1365-2966.2005.08881.x](https://doi.org/10.1111/j.1365-2966.2005.08881.x)
- Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, *ApJ*, 237, 692, doi: [10.1086/157917](https://doi.org/10.1086/157917)
- Le Digabel, S. 2011, *ACM Trans. Math. Softw.*, 37, doi: [10.1145/1916461.1916468](https://doi.org/10.1145/1916461.1916468)
- Li, J., Liu, C., Zhang, Z.-Y., et al. 2023, *Nature*, 613, 460, doi: [10.1038/s41586-022-05488-1](https://doi.org/10.1038/s41586-022-05488-1)
- Liepold, C. M., Quenneville, M. E., Ma, C.-P., et al. 2020, *ApJ*, 891, 4, doi: [10.3847/1538-4357/ab6f71](https://doi.org/10.3847/1538-4357/ab6f71)
- Lin, D. N. C., & Faber, S. M. 1983, *ApJL*, 266, L21, doi: [10.1086/183971](https://doi.org/10.1086/183971)
- Lipka, M., & Thomas, J. 2021, *MNRAS*, 504, 4599, doi: [10.1093/mnras/stab1092](https://doi.org/10.1093/mnras/stab1092)
- Lisker, T., Glatt, K., Westera, P., & Grebel, E. K. 2006a, *AJ*, 132, 2432, doi: [10.1086/508414](https://doi.org/10.1086/508414)
- Lisker, T., Grebel, E. K., & Binggeli, B. 2006b, *AJ*, 132, 497, doi: [10.1086/505045](https://doi.org/10.1086/505045)
- Lisker, T., Grebel, E. K., Binggeli, B., & Glatt, K. 2007, *ApJ*, 660, 1186, doi: [10.1086/513090](https://doi.org/10.1086/513090)
- Lokas, E. L., Kazantzidis, S., Klimentowski, J., Mayer, L., & Callegari, S. 2010, *ApJ*, 708, 1032, doi: [10.1088/0004-637X/708/2/1032](https://doi.org/10.1088/0004-637X/708/2/1032)
- Magorrian, J. 1999, *MNRAS*, 302, 530, doi: [10.1046/j.1365-8711.1999.02135.x](https://doi.org/10.1046/j.1365-8711.1999.02135.x)

- Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, *A&A*, 570, A13, doi: [10.1051/0004-6361/201423496](https://doi.org/10.1051/0004-6361/201423496)
- Mancera Piña, P. E., Posti, L., Fraternali, F., Adams, E. A. K., & Oosterloo, T. 2021, *A&A*, 647, A76, doi: [10.1051/0004-6361/202039340](https://doi.org/10.1051/0004-6361/202039340)
- Maraston, C. 1998, *MNRAS*, 300, 872, doi: [10.1046/j.1365-8711.1998.01947.x](https://doi.org/10.1046/j.1365-8711.1998.01947.x)
- . 2005, *MNRAS*, 362, 799, doi: [10.1111/j.1365-2966.2005.09270.x](https://doi.org/10.1111/j.1365-2966.2005.09270.x)
- Martín-Navarro, I., Vazdekis, A., La Barbera, F., et al. 2015, *ApJL*, 806, L31, doi: [10.1088/2041-8205/806/2/L31](https://doi.org/10.1088/2041-8205/806/2/L31)
- Martín-Navarro, I., Spiniello, C., Tortora, C., et al. 2023, *MNRAS*, 521, 1408, doi: [10.1093/mnras/stad503](https://doi.org/10.1093/mnras/stad503)
- Martínez-García, A. M., del Pino, A., Aparicio, A., van der Marel, R. P., & Watkins, L. L. 2021, *MNRAS*, 505, 5884, doi: [10.1093/mnras/stab1568](https://doi.org/10.1093/mnras/stab1568)
- Mateo, M. L. 1998, *ARA&A*, 36, 435, doi: [10.1146/annurev.astro.36.1.435](https://doi.org/10.1146/annurev.astro.36.1.435)
- Mayer, L., Governato, F., Colpi, M., et al. 2001, *ApJ*, 559, 754, doi: [10.1086/322356](https://doi.org/10.1086/322356)
- McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, *MNRAS*, 448, 3484, doi: [10.1093/mnras/stv105](https://doi.org/10.1093/mnras/stv105)
- Mehrgan, K., Thomas, J., Saglia, R., et al. 2019, *ApJ*, 887, 195, doi: [10.3847/1538-4357/ab5856](https://doi.org/10.3847/1538-4357/ab5856)
- Mehrgan, K., Thomas, J., Saglia, R., Parikh, T., & Bender, R. 2023, *ApJ*, 948, 79, doi: [10.3847/1538-4357/acbf2e](https://doi.org/10.3847/1538-4357/acbf2e)
- Mehrgan, K., Thomas, J., Saglia, R., et al. 2024, *ApJ*, 961, 127, doi: [10.3847/1538-4357/acfe09](https://doi.org/10.3847/1538-4357/acfe09)
- Michielsen, D., Boselli, A., Conselice, C. J., et al. 2008, *MNRAS*, 385, 1374, doi: [10.1111/j.1365-2966.2008.12846.x](https://doi.org/10.1111/j.1365-2966.2008.12846.x)
- Moore, B., Governato, F., Quinn, T., Stadel, J., & Lake, G. 1998, *ApJL*, 499, L5, doi: [10.1086/311333](https://doi.org/10.1086/311333)
- Napolitano, N. R., Romanowsky, A. J., & Tortora, C. 2010, *MNRAS*, 405, 2351, doi: [10.1111/j.1365-2966.2010.16710.x](https://doi.org/10.1111/j.1365-2966.2010.16710.x)
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 475, 624, doi: [10.1093/mnras/stx3040](https://doi.org/10.1093/mnras/stx3040)
- Neureiter, B., de Nicola, S., Thomas, J., et al. 2023a, *MNRAS*, 519, 2004, doi: [10.1093/mnras/stac3652](https://doi.org/10.1093/mnras/stac3652)
- Neureiter, B., Thomas, J., Rantala, A., et al. 2023b, *ApJ*, 950, 15, doi: [10.3847/1538-4357/acffa](https://doi.org/10.3847/1538-4357/acffa)
- Ocvirk, P., Pichon, C., Lançon, A., & Thiébaud, E. 2006, *MNRAS*, 365, 74, doi: [10.1111/j.1365-2966.2005.09323.x](https://doi.org/10.1111/j.1365-2966.2005.09323.x)
- Ohlson, D., Seth, A. C., Gallo, E., Baldassare, V. F., & Greene, J. E. 2024, *AJ*, 167, 31, doi: [10.3847/1538-3881/acf7bc](https://doi.org/10.3847/1538-3881/acf7bc)
- Oldham, L., & Auger, M. 2018, *MNRAS*, 474, 4169, doi: [10.1093/mnras/stx2969](https://doi.org/10.1093/mnras/stx2969)
- Parikh, T., Saglia, R., Thomas, J., et al. 2024, arXiv e-prints, arXiv:2402.06628, doi: [10.48550/arXiv.2402.06628](https://doi.org/10.48550/arXiv.2402.06628)
- Parikh, T., Thomas, D., Maraston, C., et al. 2018, *MNRAS*, 477, 3954, doi: [10.1093/mnras/sty785](https://doi.org/10.1093/mnras/sty785)
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *A&A*, 412, 45, doi: [10.1051/0004-6361:20031411](https://doi.org/10.1051/0004-6361:20031411)
- Paudel, S., Lisker, T., & Kuntschner, H. 2011, *MNRAS*, 413, 1764, doi: [10.1111/j.1365-2966.2011.18256.x](https://doi.org/10.1111/j.1365-2966.2011.18256.x)
- Paudel, S., Lisker, T., Kuntschner, H., Grebel, E. K., & Glatt, K. 2010, *MNRAS*, 405, 800, doi: [10.1111/j.1365-2966.2010.16507.x](https://doi.org/10.1111/j.1365-2966.2010.16507.x)
- Paudel, S., & Ree, C. H. 2014, *ApJL*, 796, L14, doi: [10.1088/2041-8205/796/1/L14](https://doi.org/10.1088/2041-8205/796/1/L14)
- Paudel, S., Yoon, S.-J., Yoo, J., et al. 2023, *ApJS*, 265, 57, doi: [10.3847/1538-4365/acbfa7](https://doi.org/10.3847/1538-4365/acbfa7)
- Peebles, P. J. E. 1969, *ApJ*, 155, 393, doi: [10.1086/149876](https://doi.org/10.1086/149876)
- Poci, A., McDermid, R. M., Zhu, L., & van de Ven, G. 2019, *MNRAS*, 487, 3776, doi: [10.1093/mnras/stz1154](https://doi.org/10.1093/mnras/stz1154)
- Posacki, S., Cappellari, M., Treu, T., Pellegrini, S., & Ciotti, L. 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 493, doi: [10.1093/mnras/stu2098](https://doi.org/10.1093/mnras/stu2098)
- Posti, L., Fraternali, F., Di Teodoro, E. M., & Pezzulli, G. 2018, *A&A*, 612, L6, doi: [10.1051/0004-6361/201833091](https://doi.org/10.1051/0004-6361/201833091)
- Prugniel, P., Koleva, M., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2007a, in *Stellar Populations as Building Blocks of Galaxies*, ed. A. Vazdekis & R. Peletier, Vol. 241, 68–72, doi: [10.1017/S1743921307007454](https://doi.org/10.1017/S1743921307007454)
- Prugniel, P., & Soubiran, C. 2001, *A&A*, 369, 1048, doi: [10.1051/0004-6361:20010163](https://doi.org/10.1051/0004-6361:20010163)
- Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007b, arXiv e-prints, astro, <https://arxiv.org/abs/astro-ph/0703658>
- Pulsoni, C., Gerhard, O., Fall, S. M., et al. 2023, *A&A*, 674, A96, doi: [10.1051/0004-6361/202346234](https://doi.org/10.1051/0004-6361/202346234)
- Quenneville, M. E., Liepold, C. M., & Ma, C.-P. 2022, *ApJ*, 926, 30, doi: [10.3847/1538-4357/ac3e68](https://doi.org/10.3847/1538-4357/ac3e68)
- Rantala, A., Johansson, P. H., Naab, T., Thomas, J., & Frigo, M. 2018, *ApJ*, 864, 113, doi: [10.3847/1538-4357/aada47](https://doi.org/10.3847/1538-4357/aada47)
- Rix, H.-W., Franx, M., Fisher, D., & Illingworth, G. 1992, *ApJL*, 400, L5, doi: [10.1086/186635](https://doi.org/10.1086/186635)
- Romanowsky, A. J., & Fall, S. M. 2012, *ApJS*, 203, 17, doi: [10.1088/0067-0049/203/2/17](https://doi.org/10.1088/0067-0049/203/2/17)
- Romero-Gómez, J., Peletier, R. F., Aguerri, J. A. L., & Smith, R. 2024, arXiv e-prints, arXiv:2404.15519, doi: [10.48550/arXiv.2404.15519](https://doi.org/10.48550/arXiv.2404.15519)

- Romero-Gómez, J., Aguerri, J. A. L., Peletier, R. F., et al. 2023a, *Monthly Notices of the Royal Astronomical Society*, 527, 9715, doi: [10.1093/mnras/stad3801](https://doi.org/10.1093/mnras/stad3801)
- Romero-Gómez, J., Peletier, R. F., Aguerri, J. A. L., et al. 2023b, *Monthly Notices of the Royal Astronomical Society*, 522, 130, doi: [10.1093/mnras/stad953](https://doi.org/10.1093/mnras/stad953)
- Rubin, V. C., Graham, J. A., & Kenney, J. D. P. 1992, *ApJL*, 394, L9, doi: [10.1086/186460](https://doi.org/10.1086/186460)
- Rybicki, G. B. 1987, in *IAU Symposium*, Vol. 127, *Structure and Dynamics of Elliptical Galaxies*, ed. P. T. de Zeeuw, 397, doi: [10.1007/978-94-009-3971-4_41](https://doi.org/10.1007/978-94-009-3971-4_41)
- Ryś, A., Falcón-Barroso, J., & van de Ven, G. 2013, *MNRAS*, 428, 2980, doi: [10.1093/mnras/sts245](https://doi.org/10.1093/mnras/sts245)
- Ryś, A., Koleva, M., Falcón-Barroso, J., et al. 2015, *MNRAS*, 452, 1888, doi: [10.1093/mnras/stv1364](https://doi.org/10.1093/mnras/stv1364)
- Salpeter, E. E. 1955, *ApJ*, 121, 161, doi: [10.1086/145971](https://doi.org/10.1086/145971)
- Sandage, A., Binggeli, B., & Tammann, G. A. 1985, *AJ*, 90, 1759, doi: [10.1086/113875](https://doi.org/10.1086/113875)
- Santucci, G., Brough, S., van de Sande, J., et al. 2022, *ApJ*, 930, 153, doi: [10.3847/1538-4357/ac5bd5](https://doi.org/10.3847/1538-4357/ac5bd5)
- . 2023, *MNRAS*, 521, 2671, doi: [10.1093/mnras/stad713](https://doi.org/10.1093/mnras/stad713)
- Schechter, P. L., & Gunn, J. E. 1978, *AJ*, 83, 1360, doi: [10.1086/112324](https://doi.org/10.1086/112324)
- Schwarzschild, M. 1979, *ApJ*, 232, 236, doi: [10.1086/157282](https://doi.org/10.1086/157282)
- Scorza, C., & Bender, R. 1995, *A&A*, 293, 20
- Scott, N., Eftekhari, F. S., Peletier, R. F., et al. 2020, *MNRAS*, 497, 1571, doi: [10.1093/mnras/staa2042](https://doi.org/10.1093/mnras/staa2042)
- Seo, M., & Ann, H. B. 2023, *MNRAS*, 520, 5521, doi: [10.1093/mnras/stad425](https://doi.org/10.1093/mnras/stad425)
- Simien, F., & Prugniel, P. 2002, *A&A*, 384, 371, doi: [10.1051/0004-6361:20020071](https://doi.org/10.1051/0004-6361:20020071)
- Sirianni, M., Jee, M. J., Benítez, N., et al. 2005, *PASP*, 117, 1049, doi: [10.1086/444553](https://doi.org/10.1086/444553)
- Smith, R. J., Lucey, J. R., & Carter, D. 2012, *MNRAS*, 426, 2994, doi: [10.1111/j.1365-2966.2012.21922.x](https://doi.org/10.1111/j.1365-2966.2012.21922.x)
- Spavone, M., Iodice, E., D'Ago, G., et al. 2022, *A&A*, 663, A135, doi: [10.1051/0004-6361/202243290](https://doi.org/10.1051/0004-6361/202243290)
- Sybiliska, A., Lisker, T., Kuntschner, H., et al. 2017, *MNRAS*, 470, 815, doi: [10.1093/mnras/stx1138](https://doi.org/10.1093/mnras/stx1138)
- Thomas, D., Maraston, C., & Bender, R. 2003, *MNRAS*, 339, 897, doi: [10.1046/j.1365-8711.2003.06248.x](https://doi.org/10.1046/j.1365-8711.2003.06248.x)
- Thomas, D., Maraston, C., & Johansson, J. 2011a, *MNRAS*, 412, 2183, doi: [10.1111/j.1365-2966.2010.18049.x](https://doi.org/10.1111/j.1365-2966.2010.18049.x)
- Thomas, J., & Lipka, M. 2022, *MNRAS*, 514, 6203, doi: [10.1093/mnras/stac1581](https://doi.org/10.1093/mnras/stac1581)
- Thomas, J., Ma, C.-P., McConnell, N. J., et al. 2016, *Nature*, 532, 340, doi: [10.1038/nature17197](https://doi.org/10.1038/nature17197)
- Thomas, J., Saglia, R. P., Bender, R., Erwin, P., & Fabricius, M. 2014, *ApJ*, 782, 39, doi: [10.1088/0004-637X/782/1/39](https://doi.org/10.1088/0004-637X/782/1/39)
- Thomas, J., Saglia, R. P., Bender, R., et al. 2007, *MNRAS*, 382, 657, doi: [10.1111/j.1365-2966.2007.12434.x](https://doi.org/10.1111/j.1365-2966.2007.12434.x)
- . 2009a, *ApJ*, 691, 770, doi: [10.1088/0004-637X/691/1/770](https://doi.org/10.1088/0004-637X/691/1/770)
- . 2004, *MNRAS*, 353, 391, doi: [10.1111/j.1365-2966.2004.08072.x](https://doi.org/10.1111/j.1365-2966.2004.08072.x)
- Thomas, J., Jesseit, R., Saglia, R. P., et al. 2009b, *MNRAS*, 393, 641, doi: [10.1111/j.1365-2966.2008.14238.x](https://doi.org/10.1111/j.1365-2966.2008.14238.x)
- Thomas, J., Saglia, R. P., Bender, R., et al. 2011b, *MNRAS*, 415, 545, doi: [10.1111/j.1365-2966.2011.18725.x](https://doi.org/10.1111/j.1365-2966.2011.18725.x)
- Toloba, E., Boselli, A., Cenarro, A. J., et al. 2011, *A&A*, 526, A114, doi: [10.1051/0004-6361/201015344](https://doi.org/10.1051/0004-6361/201015344)
- Toloba, E., Guhathakurta, P., Peletier, R. F., et al. 2014, *ApJS*, 215, 17, doi: [10.1088/0067-0049/215/2/17](https://doi.org/10.1088/0067-0049/215/2/17)
- Toloba, E., Guhathakurta, P., Boselli, A., et al. 2015, *ApJ*, 799, 172, doi: [10.1088/0004-637X/799/2/172](https://doi.org/10.1088/0004-637X/799/2/172)
- Tortora, C., La Barbera, F., & Napolitano, N. R. 2016, *MNRAS*, 455, 308, doi: [10.1093/mnras/stv2250](https://doi.org/10.1093/mnras/stv2250)
- Tortora, C., Napolitano, N. R., Saglia, R. P., et al. 2014, *MNRAS*, 445, 162, doi: [10.1093/mnras/stu1712](https://doi.org/10.1093/mnras/stu1712)
- Tortora, C., Posti, L., Koopmans, L. V. E., & Napolitano, N. R. 2019, *MNRAS*, 489, 5483, doi: [10.1093/mnras/stz2320](https://doi.org/10.1093/mnras/stz2320)
- van de Sande, J., Bland-Hawthorn, J., Brough, S., et al. 2017, *MNRAS*, 472, 1272, doi: [10.1093/mnras/stx1751](https://doi.org/10.1093/mnras/stx1751)
- van den Bosch, R. C. E., & de Zeeuw, P. T. 2010, *MNRAS*, 401, 1770, doi: [10.1111/j.1365-2966.2009.15832.x](https://doi.org/10.1111/j.1365-2966.2009.15832.x)
- van den Bosch, R. C. E., & van de Ven, G. 2009, *MNRAS*, 398, 1117, doi: [10.1111/j.1365-2966.2009.15177.x](https://doi.org/10.1111/j.1365-2966.2009.15177.x)
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525, doi: [10.1086/172534](https://doi.org/10.1086/172534)
- van Dokkum, P., Conroy, C., Villaume, A., Brodie, J., & Romanowsky, A. J. 2017, *ApJ*, 841, 68, doi: [10.3847/1538-4357/aa7135](https://doi.org/10.3847/1538-4357/aa7135)
- van Dokkum, P. G. 2008, *ApJ*, 674, 29, doi: [10.1086/525014](https://doi.org/10.1086/525014)
- van Dokkum, P. G., & Conroy, C. 2010, *Nature*, 468, 940, doi: [10.1038/nature09578](https://doi.org/10.1038/nature09578)
- van Zee, L., Skillman, E. D., & Haynes, M. P. 2004, *AJ*, 128, 121, doi: [10.1086/421368](https://doi.org/10.1086/421368)
- Walsh, J. L., van den Bosch, R. C. E., Barth, A. J., & Sarzi, M. 2012, *ApJ*, 753, 79, doi: [10.1088/0004-637X/753/1/79](https://doi.org/10.1088/0004-637X/753/1/79)
- Wang, B., Cappellari, M., Peng, Y., & Graham, M. 2020, *MNRAS*, 495, 1958, doi: [10.1093/mnras/staa1325](https://doi.org/10.1093/mnras/staa1325)
- Williams, T. B., & Schwarzschild, M. 1979, *ApJS*, 41, 209, doi: [10.1086/190616](https://doi.org/10.1086/190616)
- Yang, Y., Hammer, F., Fouquet, S., et al. 2014, *MNRAS*, 442, 2419, doi: [10.1093/mnras/stu931](https://doi.org/10.1093/mnras/stu931)

Zeng, G., Wang, L., Gao, L., & Yang, H. 2024, arXiv e-prints, arXiv:2404.14184, doi: [10.48550/arXiv.2404.14184](https://doi.org/10.48550/arXiv.2404.14184)

Zhao, H. 1996, MNRAS, 278, 488, doi: [10.1093/mnras/278.2.488](https://doi.org/10.1093/mnras/278.2.488)

Zhu, K., Lu, S., Cappellari, M., et al. 2024, MNRAS, 527, 706, doi: [10.1093/mnras/stad3213](https://doi.org/10.1093/mnras/stad3213)

Zhu, L., van den Bosch, R., van de Ven, G., et al. 2018, MNRAS, 473, 3000, doi: [10.1093/mnras/stx2409](https://doi.org/10.1093/mnras/stx2409)

Chapter 4

The VIRUS-dE Survey II: Cuspy and round halos in dwarf ellipticals - A result of early assembly?

Bibliographic information

Lipka Mathias, Jens Thomas, Roberto Saglia, Ralf Bender, Maximilian Fabricius, Christian Partmann - The VIRUS-dE Survey II: Cuspy and round halos in dwarf ellipticals - A result of early assembly?, *The Astrophysical Journal*, Volume TBA, Issue TBA, pp.???-???
DOI: TBA

The Author's contribution

As first author, I wrote all sections and produced all figures in the manuscript. Jens Thomas provided extensive support to this work in the form of numerous discussions, suggestions, and text revisions. Roberto Saglia and Ralf Bender provided further corrections and important suggestions to improve the paper. Christian Partmann provided me with the N-body simulation that I dynamically modelled using the LIBMOD code provided by Jens Thomas. All authors provided feedback and corrections for the final version of the manuscript.

Copyright

©2024 MNRAS
reprinted on pages 124- 169

The VIRUS-dE Survey II: Cuspy and round halos in dwarf ellipticals - A result of early assembly?

MATHIAS LIPKA ^{1,2} JENS THOMAS ^{1,2} ROBERTO SAGLIA ^{1,2} RALF BENDER ^{1,2} MAXIMILIAN FABRICIUS ^{1,2} AND
CHRISTIAN PARTMANN ³

¹Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching

²Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany

³Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748, Garching, Germany

Submitted to ApJ

ABSTRACT

We analyze the dark matter (DM) halos of a sample of dwarf Ellipticals (dE) and discuss cosmological and evolutionary implications. Using orbit modeling we recover their density slopes and, for the first time, the halo flattening. We find the ‘cusp-core’ tension is mild, on average dEs have central slopes slightly below the Navarro–Frenk–White (NFW) predictions. However, the measured flattenings are still more spherical than cosmological simulations predict. Unlike brighter ETGs the total density slopes of dEs are shallower, and their average DM density does not follow their scaling relation with luminosity. Conversely, dE halos are denser and the densities steeper than in LTGs. We find average DM density and slope are strongly correlated with the environment and moderately with the angular momentum. Central, non-rotating dEs have dense and cuspy halos, whereas rotating dEs in Virgo’s outskirts are more cored and less dense. This can be explained by a delayed formation of the dEs in the cluster outskirts, or alternatively, by the accumulated baryonic feedback the dEs in the outskirts have experienced during their very different star formation history. Our results suggest halo profiles are not universal (they depend on assembly conditions) and they evolve only mildly due to internal feedback. We conclude dEs in the local Universe have assembled at a higher redshift than local spirals. In these extreme conditions (e.g. star-formation, halo assembly) were very different, suggesting no new dEs are formed at present.

Keywords: Galaxy structure(622) — Galaxy formation(595) — Dwarf elliptical galaxies(415) — Virgo Cluster(1772) — Galaxy dark matter halos(1880) — Dark matter distribution(356)

1. INTRODUCTION

In the standard cosmological model (Λ CDM) structures like galaxies have assembled from collapsing dark matter (DM) over-densities as baryons followed them to build the galaxies observed today. In the hierarchical formation scenario the *largest* galaxies are thought to have formed via mergers of *smaller* galaxies, i.e. smaller DM halos, which has dramatically changed their mass and kinematic structure. The smaller dwarf galaxies on the other hand, which avoided merging to this day,

should have pristine halos which makes them an ideal probe of the initial DM over-density collapse.

In this framework, dwarf ellipticals (dEs) that inhabit an *intermediate* mass regime ($\log_{10}(M_*/M_\odot) \approx 7 - 9$) are often thought to be the largest *fundamental* building blocks within the sequence of the quiescent early-type galaxies (ETGs) that have not formed via merging. However, while they may have avoided merging to this day, they are also much more exposed to their environment due to their much shallower potential wells. Over time *internal* feedback processes like star-formation, and/or *environmental* processes like ram-pressure stripping RMS (Gunn & Gott 1972; Lin & Faber 1983) and galaxy harassment (Moore et al. 1998), could have modified the distribution of baryons and as such also that of

the DM. Therefore the present-day structure of dE dark matter halos may not only be an excellent probe of the underlying cosmology but also an avenue to investigate the effect cluster environments and/or internal feedback has had on the dark matter as time passed.

While overall quite successful in explaining the observed clustering of mass on large scales (Croft et al. 2002; Spergel et al. 2003; Springel et al. 2006), Λ CDM predictions are much harder to reconcile with observational evidence on the smaller galaxy scales. Firstly, the statistical occurrence and distribution of the dwarf galaxies in the local universe differ from cosmological predictions, which is known as the ‘missing satellite’ and the ‘too big to fail’-problem (e.g. Klypin et al. 1999; Boylan-Kolchin et al. 2011, 2012). Secondly, observational constraints on the individual DM halo distributions suggest they are closer to spherical than Λ CDM simulations anticipated (e.g. Allgood et al. 2006; Hayashi et al. 2007; Chua et al. 2019; Bovy et al. 2016; Wegg et al. 2019). And thirdly, the steepness of the inner DM distribution conflicts with predictions, which is known as the ‘cusp-core problem’ and was first reported by Moore (1994) (For a review see de Blok 2010; Del Popolo & Le Delliou 2021). Simulations of halo formation suggest *cuspy* central density profiles for the halos of dwarf galaxies. Examples are the Navarro–Frenk–White profile (NFW), which has a central logarithmic slope of -1 , or even cuspier halos with slopes of ~ -1.5 (e.g. Diemand et al. 2004; Diemand & Moore 2011; Moore et al. 1998, 2001; Klypin et al. 2001, 2011; Navarro et al. 2010). This stands in contrast to the majority of observational findings, which often find cored density distributions. For example, the rotation curve modeling of HI disks suggests that at least the smaller dwarf galaxies have a strong preference towards *cored* halos (e.g. de Blok & Bosma 2002; de Blok et al. 2008; Donato et al. 2009; Oh et al. 2011b; Plana et al. 2010).

If these discrepancies between Λ CDM predictions and observations are quantified accurately, then one may be able to identify the reason behind it: be it an exotic nature of dark matter particles (Spergel & Steinhardt 2000; Marsh & Silk 2014; Elbert et al. 2015) or the feedback of baryonic physics on the dark matter (de Souza et al. 2011; Gnedin & Zhao 2002; Governato et al. 2010; Madau et al. 2014; Navarro et al. 1996a; Oh et al. 2011a). However, the majority of observational evidence for cored halos stems from dynamical modeling using gas as tracer of the gravitational potential and is thus mostly restricted to late-type galaxies. Observational constraints on the degree of sphericity of halos are even more scarce, and the majority of evidence comes from Milky Way studies.

For *early-type* galaxies, i.e. for galaxies without significant amounts of gas, stellar-based dynamical models can be employed to infer the structure of their DM halos. However, the existing literature is mostly restricted to the very small but near-by dwarf spheroidals (dSphs) within the Local Group or to very massive ETGs ($\log_{10}(M_*/M_\odot) \gtrsim 10$) that inhabit the more distant massive galaxy clusters (Coma, Virgo, Fornax ...). In contrast, the DM structure of dEs (i.e. the intermediate mass regime) is scarcely probed even though they are by far the most common type of galaxy found in the nearby clusters. This deficiency of observational constraints in the dE regime is mostly because the Local Group only has a few dEs, while the dEs in the nearby clusters are faint and require a very high spectral resolution to be analyzed using stellar dynamical models.

This paper is part of a series aimed at studying the mass distribution, stellar populations and dynamical composition by analyzing a sample of 9 such dEs in the Virgo cluster. The first paper, Lipka et al. (2024), in the following VW–I, is a comprehensive analysis of the *stellar* structure of the dEs. There we also discuss basic properties of the dE sample, all the data sets we obtained, and describe the *dynamical* and population modeling techniques we employed to infer the intrinsic 3D structure. The current paper is focused on the dark matter structure and its interpretation in the broader cosmological and galaxy evolutionary context. It is aimed to fill the gap in our understanding of dark matter in the intermediate mass regime of early-type galaxies.

The current paper is organized as follows: In Section 2 we briefly describe the dE sample we obtained and recapitulate some of the main findings of VW–I. In Section 3 we explain how we specifically modified our dynamical modeling technique to optimally recover the 3D density distributions of the dark matter halos. In preparation for this study we stress-tested this modeling approach by applying it to an N-body simulation, the results of which are discussed in App. A. Section 4 shows the dynamical constraints and recovered dark mass distributions of the dEs sample. Using these modeling results we then examine whether the slopes and flattening of the halos of dEs are in tension with Λ CDM predictions or not (Section 5). Under the umbrella of the Λ CDM paradigm we then discuss what our results imply regarding the formation and evolution of dEs and how they are related to other galaxy types (Section 6). The paper concludes with a summary in Section 7.

2. THE VIRUS-W DWARF SAMPLE

The sample we analyze in this paper consists of 9 dEs with stellar masses $\log_{10}(M_*) \in [8.5, 9.5]$ which inhabit

different environments within the Virgo cluster, ranging from its center to just beyond its virial radius. The basic properties (like distance) of the galaxies we adopted for the dynamical modeling can be found in Table 2 of VW–I. To distinguish the individual galaxies in our sample, we keep the same color-coding we used in VW–I. The VCC-Catalog ID of each galaxy and its corresponding color can be inferred from Fig. 1 or Tab. 1 which shows some of the most important quantities we measured in this paper.

The main data set we obtained with the integral-field-unit (IFU) spectrograph VIRUS-W (Fabricius et al. 2008, 2012) at the Harlan J. Smith Telescope (McDonald Observatory). VIRUS-W’s very high spectral resolution ($R = 7900$ to 9000) is essential to measure the low velocity dispersions of the stars in dEs, because spectrographs with a lower resolution tend to overestimate σ by a significant margin (cf. VW–I) which would inevitably bias any dynamical mass reconstruction. As one of the first IFU studies of dEs the VIRUS-W data allow us to access their full 2D spatially-resolved on-sky kinematic which vastly improves the constraints of the dynamical modeling required to infer the 3D density distribution of the dark matter halos analyzed here.

To obtain the line-of-sight-velocity-distributions (LOSVDs) of the stars we employed the spectral-fitting code WINGFIT (Thomas et al. in prep.) which allows us to retrieve the full extent of the information contained in the spectra well beyond just the mean velocity and dispersion. Higher moment information is crucial to break the mass-anisotropy degeneracy (e.g. Merrifield & Kent 1990; van der Marel & Franx 1993) and conversely enable a robust recovery of the dark matter distribution. To ensure a robust recovery of the LOSVDs we binned the spectra with the Voronoi tessellation method (Cappellari & Copin 2003) and excluded any bins that did not fulfil our S/N requirement. The remaining data covers the kinematic out to approximately 1 effective radius.

For a detailed description of the data reduction/preparation and the final resulting LOSVDs we refer to VW–I, in it we also show our results from stellar population modeling and the *stellar* density reconstruction we obtained from our dynamical modeling implementation (which is also used here). In short we find: stellar populations are spatially homogeneous, but display a larger variety in age with some having stopped forming stars only recently while others did so 12 Gyrs ago. Dynamically, the dEs have a more isotropic orbit structure and a suppressed angular momentum compared to other galaxy types (see also Scott et al. 2020). We find that their mass-to-light ratios are

anti-correlated with their single stellar population (SSP) age and we identify two possible explanations: i) In the Virgo cluster dEs have been formed continuously starting 12 Gyrs ago until now and their initial mass function (IMF) changed with their formation epoch, ii) or the bulk of Virgo dEs has formed early on in the same epoch, but (subject to internal/external influences) have experienced varying degrees of extended star formation history (SFH) with some dE being quenched shortly after gravitational collapse while others were able to sustain several Gyrs of continuous or bursty star formation until they were eventually quenched. These processes could have left imprints in the distribution of the dark matter, which we will investigate in this work.

3. RECOVERING DARK MATTER WITH DYNAMICAL MODELS

The modeling code we employ is a state-of-the-art axisymmetric implementation of the Schwarzschild orbit superposition technique (Schwarzschild 1979; Thomas et al. 2004). In short, the modeling principle is simple: For a given galaxy a number of *candidate* mass models is established, and then a set of representative orbits in each of the corresponding potentials is integrated. Each of the orbits is given a weight, and the weighted superposition determines the stellar phase-space density of the orbit model. Given a set of observed data, e.g. LOSVDs, an optimal set of orbit weights can be determined by fitting a candidate model’s LOSVDs to the data¹. To find the candidate mass model that best represents a given galaxy, the fit of each candidate model to the data is then compared using an evaluation statistic. A popular choice for this statistic is χ^2 , however, in Lipka & Thomas (2021) we demonstrated that χ^2 is biased due to the varying fit flexibility of different candidate models. We developed a new model selection approach (Thomas & Lipka 2022) that takes this flexibility into account in its evaluation. The corresponding evaluation statistic is an extension of the Akaike information (Akaike 1973, 1974) criterion and called AIC_p . Analogous to a χ^2 approach, the mass model with the minimum AIC_p is deemed to be the best representation of the galaxy under investigation. Beyond the AIC_p which improves modeling constraints in general, we made some adjustments in the setup of the candidate models and the analysis of results with the specific goal to ensure an

¹ This involves a regularization to avoid overfitting since the number of weights is typically larger than the number of data constraints. In Lipka & Thomas (2021) and Thomas & Lipka (2022) we introduced a novel data-driven approach that allows the determination of the optimum amount of regularization to avoid both over- and under-fitting.

Table 1. Table with some of the important quantities we measured. From left to right: The average DM density within $1r_e$ in M_\odot/kpc^3 , the mass-weighted slope (eq. 6) of the total density, the volume averaged slope η_{DM} within 0.8kpc (eq. 4), the axis ratio of the halo, the (mean) axis ratio of stars, the dark matter fraction within $1r_e$, the SSP age in Gyrs, metallicity, [Mg/Fe] ratio at $r = 2.5''$, stellar angular momentum in [$\text{kpc} \cdot \text{km/s}$], the angular momentum parameter. Total stellar mass and distance to M87 are tabulated in VW–I.

VCC ID	Color	$\log_{10}(\overline{\rho_{DM}})$	γ_{MW}	η_{DM}	q_*	q_{DM}	f_{DM}	Age [Gyr]	[Z/H]	[Mg/Fe]	$\log_{10}(j_*)$	$\lambda_{e/2}$
VCC 200		7.324	-1.944	-1.113	0.895	0.9	0.33	11.2	-0.63	0.28	0.957	0.208
VCC 308		7.295	-1.540	-0.521	0.826	1.0	0.24	2.8	-0.24	0.08	1.014	0.222
VCC 543		7.138	-1.791	-0.715	0.612	0.9	0.17	6.3	-0.34	0.21	1.488	0.375
VCC 856		7.724	-1.282	-0.713	0.609	1.0	0.56	7.8	-0.49	0.23	1.719	0.330
VCC 1261		7.756	-1.660	-1.426	0.713	1.0	0.53	6.6	-0.30	0.16	1.085	0.046
VCC 1528		8.048	-1.815	-1.414	0.839	0.7	0.35	6.7	-0.28	0.19	0.307	0.028
VCC 1861		7.616	-1.388	-0.719	0.897	0.9	0.54	10.0	-0.24	0.18	1.176	0.109
VCC 1910		7.591	-2.067	-1.111	0.865	1.0	0.25	2.0	0.35	0.19	1.020	0.126
VCC 2048		7.199	-1.805	-0.421	0.423	0.5	0.07	3.5	-0.20	0.17	1.366	0.256

unbiased and accurate recovery of the dark matter. In the following section we discuss those modifications. A stress-test of this entire modeling procedure applied to an N-body simulation that is placed under similar conditions as the VIRUS-W dE observations can be found in App. A.

3.1. A flexible halo mass model

The choice which candidate mass models are being probed in the first place lies in the modellers choice, but it is crucial as it may distort the dynamical constraints. We plan to investigate this on a general methodical level in Lipka et al. (in prep.). In the following this paper will be referenced as LT, but it is not part of the VIRUS-W dE survey as it is not concerned with dEs in particular.

For early-type galaxies where gas contribution is negligible, it is common to describe the models with a 3-component density distribution:

$$\rho(r) = \Upsilon_* \cdot \nu + \rho_{DM} + M_\bullet \cdot \delta(r) \quad (1)$$

Here the stellar component is determined by the stellar mass-to-light ratio Υ_* and the 3D luminosity distribution ν . The latter is obtained from the deprojection of the observed photometry (see VW–I). The mass M_\bullet of the supermassive central black hole (SMBH) and the dark matter halo density ρ_{DM} form the non-visible components of the system.

We also followed this 3 component approach for the modeling of the dE sample. However, for our dE sample an extension of e.g. the $M_\bullet - \sigma$ relation (Ferrarese & Merritt 2000; Gebhardt et al. 2000; Hu 2008) suggest black hole masses with a sphere of influence that are well below the spatial resolution of our data which would imply the black holes are undetectable. Nevertheless, in the very-low mass regime of spheroidal galaxies

some observations (Bustamante-Rosell et al. 2021) and simulations (Weller et al. 2023) suggest the possibility of a presence of ‘over-massive’ black holes that significantly exceed typical relations. Stripping of baryonic mass could move galaxies above the $M_\bullet - M_*$ relation (Pacucci et al. 2023). The degree to which such effects play a role in the more massive dEs is yet to be determined as little is known about black holes in this galaxy regime. Therefore we still decided to equip our candidate models with a variable black hole mass, even though we expect to only find an upper limit.

The stellar densities we probed allow for different (axisymmetric) *flattenings* and a *radially variable* stellar mass-to-light ratio Υ_* . We achieve this by sampling different viewing angles in the deprojection, and probing models with different inner and outer mass-to-light ratios (Υ_i at radius $r = r_i$ and Υ_o at radius $r = r_o$). Within one effective radius we find a sample mean of the stellar mass-to-light ratio Υ_* of $\sim 1.3 \pm 0.4$ (z -band), a value fairly consistent with color-based stellar mass-to-light estimates for dEs (e.g. Eftekhari et al. 2022). For details on the implementation of the stellar mass-to-light ratio gradients and a detailed analysis of the stellar mass-to-light results see VW–I.

The halo density ρ_{DM} is most susceptible to the modeler’s choice as it is usually obtained by probing a parametric description motivated by simulation results such as for example a Navarro-Frenk-White profile (Navarro et al. 1996b, 1997). In LT we will argue that it is important to adopt a description that is highly flexible. Halo parametrizations that are too restrictive in their profiles can bias the results towards specific configuration inherited from said halo model. Only a flexible halo allows one to probe a variety of different halo densities and to accurately gauge the actual constraining strength of the

dynamical models in an unbiased manner. Therefore, for the modeling of the dEs our fiducial halo model is:

$$\rho_{\text{DM}}(m, \theta) = \frac{\rho_0}{\left(\frac{m}{r_s}\right)^{\gamma_{\text{in}}} \cdot \left(1 + \frac{m}{r_s}\right)^{\gamma_{\text{out}} - \gamma_{\text{in}}}} \quad (2)$$

where m and θ are elliptical coordinates. γ_{in} and γ_{out} are the inner and outer logarithmic slopes separated by the scale radius r_s . The halo normalization ρ_0 is sampled by probing different values of $\rho_{1\text{kpc}}$ which is the density of the halo at $m = 1.0\text{kpc}$. They are related to each other by:

$$\rho_0 = \rho_{1\text{kpc}} \left(\frac{1\text{kpc}}{r_s}\right)^{\gamma_{\text{in}}} \cdot \left(1 + \frac{1\text{kpc}}{r_s}\right)^{\gamma_{\text{out}} - \gamma_{\text{in}}} \quad (3)$$

Equation 2 is essentially a Zhao-profile (Zhao 1996) in elliptical coordinates where the parameter that describes the transition width is fixed to unity. To probe the *flattening* of the halo component (independently of the stellar flattening) we additionally endowed the above halo model with a (globally constant) spheroidal flattening parameter q_{DM} .

3.2. Nuisance parameters and sampling strategy

A naive interpretation of the parameters that establish *parametric* model descriptions can be very deceiving and biased as they are often correlated with each other which makes a comparison of models with different sets of parameters non-trivial (see LT for a comprehensive discussion). For example, for our choice of the halo model the nominal values of the *asymptotic* slopes γ_{in} and γ_{out} describe the density gradients at $r \rightarrow 0$ and $r \rightarrow \infty$. But these are not actually the radii responsible for constraining the values of these two parameters. This is because they are *global* parameters, and it is the combination of γ_{in} , γ_{out} and r_s that fully determine the slope at *any* radius of the halo model. Therefore the constraints on the density slopes at every radius indirectly constrain the value of said model parameters. Instead of interpreting the halo parameters as physically meaningful parameters we treat them as *nuisance parameters* that simply serve us in setting up different trial mass distributions. Consequently, instead of interpreting and evaluating the nuisance parameters we should focus on the evaluation of the actual mass distributions they generate.

With this in mind we can also optimize the parameter sampling. The parameters in our mass model that have the strongest inter-correlations are those describing the DM halo. Inter-correlations imply that one may probe nominally very different halo parameters without actually changing the mass distribution in a meaningful way, essentially probing the same dynamical model

multiple times. Therefore to keep the available grid of candidate models efficiently small and avoid repeated sampling of essentially identical mass distributions we set up the halo parameter grid as follows: i) q_{DM} is sampled from spherical $q_{\text{DM}} = 1.0$ down to 0.6 and 0.7, thus covering the range of intrinsic flattenings observed in the respective stellar systems. Only for VCC 2048 we extended the range down to 0.3 as the AIC_p showed improvements down to $q_{\text{DM}} = 0.5$. ii) The scale radius r_s is probed between 0.1kpc and 5.0kpc, i.e. starting from a fraction of the spatial resolution out to radii several times the size of the FoV. iii) The asymptotic inner slope γ_{in} is varied between 0.0 and 1.5, whereas γ_{out} is sampled for a broader range from -1.0 to 3.0. This broadens the space of possible models, allowing even unrealistic halo distribution with a positive radial density gradient (e.g. if $r_s \rightarrow 0$ and $\gamma_{\text{out}} = -1$) or a sign reversal in the mass gradient. The goal of this sampling choice is to allow as many different mass distributions as possible, retaining generality, all the while maximizing the differences between individual candidate models as much as possible to avoid redundant sampling.

An illustration of the variety of halo models that are probed by this sampling choice is shown in Fig. 1. The profiles and shapes range from galaxies with halos much more massive than their baryonic component to galaxies with virtually no Dark matter contribution. Including the flattening q_{DM} , the space of probed Zhao halos encompasses $10^5 - 10^6$ different halos, densely covering the entire physically plausible density space ρ_{DM} .

The entire space of candidate models (eq. 1) is even larger and spans a 9D grid where each parameter (Υ_i , Υ_o , i , M_\bullet , ρ_0 , r_s , γ_{in} , γ_{out} , q_{DM}) is probed with 5–20 values. A full grid search is obviously unfeasible, which is why we search the grid by employing the Nonlinear Optimization by Mesh Adaptive Direct search NOMAD (Audet & Dennis 2006; Le Digabel 2011). We estimate the errors of all dynamically derived galaxy properties by evaluating the scatter between the best 25 AIC_p models, which is roughly equivalent to a $\Delta\text{AIC}_p \lesssim 10$ criterion². For details regarding this choice of error estimation see VW-I.

3.3. Where are dark matter halo constrained the best?

We expect the constraints on the mass models to be most robust in the regions where the data coverage is dense, conversely the mass recovery is uncertain and possibly biased at the smallest scales where we lack spatial resolution and at the largest scales where no

² In statistical modeling the models with a $\Delta\text{AIC} > 10$ are considered extremely unlikely (Burnham & Anderson 2002).

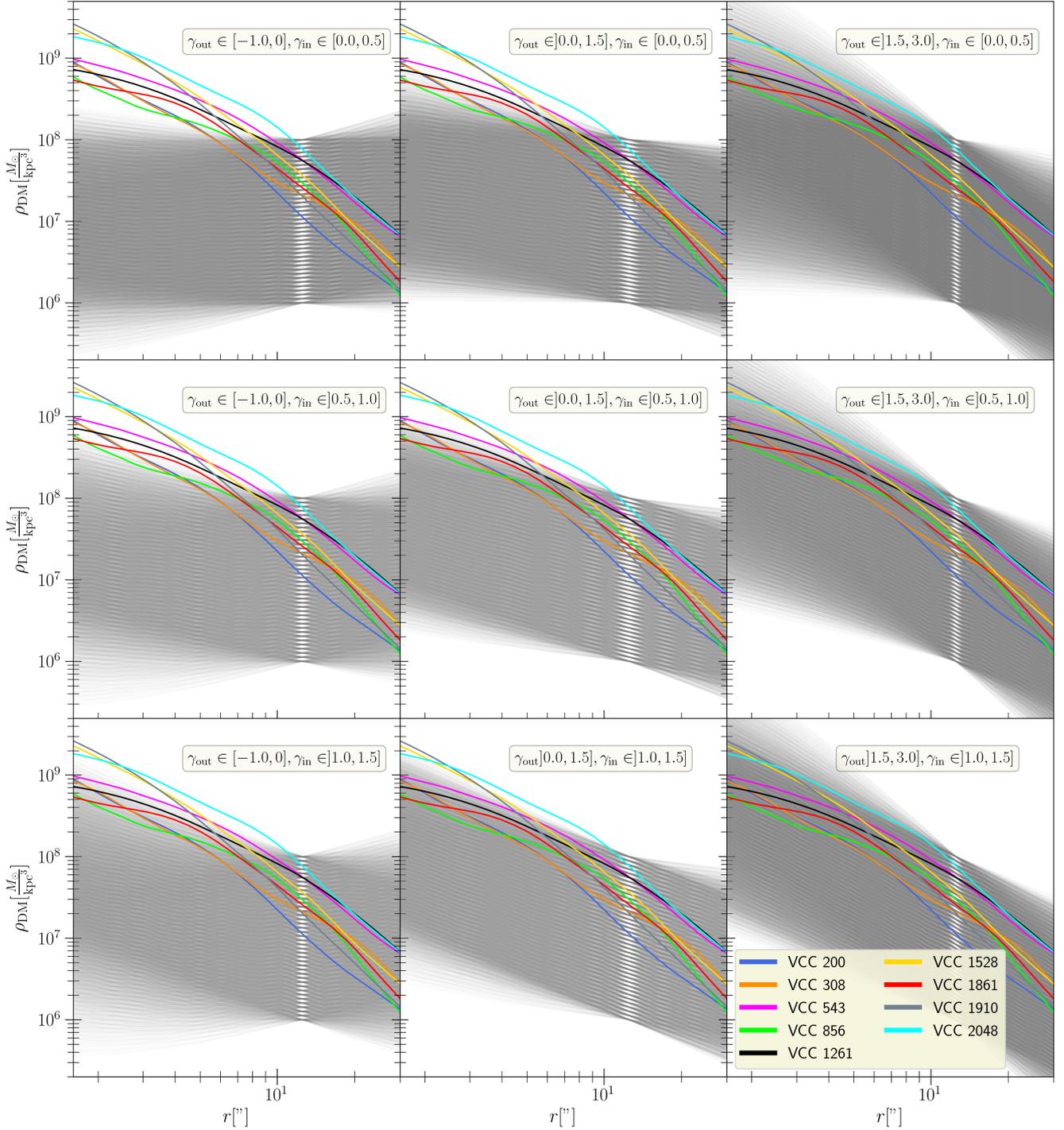


Figure 1. Representative illustration of all the candidate dark matter profiles (*gray*) we probed for each of the dEs in the sample. The color-coding for each dE we use in this study is indicated in the bottom right panel. The candidate models are obtained by sampling the Zhao-like mass description (eq. 2) on the parameter grid discussed in Sec. 3.2. In detail, the exact value of the sampling steps differed slightly from galaxy to galaxy, which is not shown here. To show the diversity in allowed profiles, we separated the candidate models into the different panels shown here according to their respective γ_{in} and γ_{out} . For simplicity we only plotted the *spherical* models ($q_{\text{DM}} = 1.0$). For comparison, we also overlay the *stellar* density distribution that we found for each of the dEs in our sample (*colored* lines).

data is available (cf. Gerhard et al. 1998; Thomas et al. 2005). This is also observed in the simulated stress-test in App. A, which is why we decided for the more cautious approach to trust and analyze the mass recovery only in the regions where data coverage is good. This implies for the dEs in our sample we should trust the mass recovery mostly in between $\sim 3''$ and $\sim 15''$ (see data coverage in VW-I), which at the distance of the Virgo cluster corresponds to $\sim 0.25\text{kpc}$ and $\sim 1.2\text{kpc}$.

This relation of constraining power and data coverage applies to the distribution of total mass in general. However, the goal of this paper is to determine the distribution of the DM halos which requires a *decomposition* of the individual mass components that make up the total mass (eq. 1). For a mass component to be dynamically detectable, this requires a significant contribution of the individual mass component to the *total* enclosed mass (see LT). For the DM component this implies the constraining power scales with the enclosed dark matter fraction f_{DM} . If the galaxy has a high dark matter fraction it means the halo may well be constrained by the data because the contribution to the total dynamical mass is an essential requirement for the models to emulate the dynamics of the whole system. In contrast, if the dark matter fraction is negligible (e.g. in the center) the *relative* uncertainty of the DM profile becomes very large. For example, one could easily double the DM density without changing the total mass and gravitational potential of the model in a dynamically detectable manner. In other words: the exact shape and profile of the halo gets very uncertain within the regions where the models suggest $f_{\text{DM}} \approx 0$.

Unfortunately for the majority of galaxies the very central parts are expected to be dominated by the baryonic/luminous matter (and/or M_{\bullet}) such that f_{DM} is a monotonically rising function eventually dominating the total mass at larger radii. If that is indeed the case for our dEs, a dynamical measurement of their *central* dark matter density and their slope becomes extremely ambiguous. Instead it is preferable to focus the analysis on the DM properties near the edge of the FoV where f_{DM} is significant and the total mass is still well constrained by the data coverage (see also LT). As discussed in VW-I, the contribution of the blue central nuclei to the central VIRUS-W bins is dynamically negligible, such that it is not worth treating them as a separate dynamical component and trying to recover their properties.

4. MODELING RESULTS

In the following we present the halo density structure that we were able to find with the dynamical modeling setup as described above. We continue with the

discussion and interpretation of these dark matter results in the subsequent sections. We discussed the *stellar* and *kinematic* structure in a detailed manner in VW-I. There we also showed how well the dynamical models are able to reproduce the observations: the spatially resolved features in the mean velocity, dispersion but also the higher-order Gauss-Hermite moments of the dEs are well reproduced by the best axisymmetric model found. The 3D deprojected luminosity density is a boundary constraint in the orbit modelling technique we use (cf. Thomas et al. 2004), i.e. it must be reproduced by each orbit model we probe. This ensures that the 2D surface brightness is reproduced as well.

4.1. Constraints on the (nuisance) parameters

Fig. 2 shows the AIC_{p} -constraints we obtained from all the orbit models that were probed on the 9D-parameter mass model grid. The stellar mass-to-light ratios, the dark matter normalization $\rho_{1\text{kpc}}$, and the inclination appear to be the most strongly constrained as indicated by sharp lower and upper boundaries in AIC_{p} . These parameters exhibit the sharpest constraints, because they directly dictate the *scale* of the global mass distribution. The asymptotic slopes and the scale radius of the halo profile show much more diversity and scatter, with some AIC_{p} constraints even reaching the edges of the explored parameter space. These parameters (r_{s} , γ_{in} , γ_{out}) are strongly inter-correlated and change the radial behavior of the mass distribution within the dynamically relevant range more indirectly.

In LT we demonstrate that, because the FoV with the kinematic constraints is always limited in radius, the values of the halo parameters themselves should not be interpreted physically. Instead these parameters should be viewed as nuisance parameters only, that allow us to construct flexible, yet smooth, halo mass distributions (cf. Fig. 1). Neither the larger scatter in AIC_{p} in these parameters nor the fact that some of these parameters can reach grid sampling limits is necessarily an issue. The halo of VCC 1261, for example, reaches the grid sampling limits at $\gamma_{\text{in}} = 0.0$ and $r_{\text{s}} = 0.1\text{kpc}$. However, an even lower r_{s} or γ_{in} would have essentially no impact on the total mass distribution at the scales where we can probe it with the models, because that would change the mass and slope predominantly at radii that are much smaller than the resolution of the VIRUS-W data ($\sim 0.13\text{kpc}$). In that case the inner slope $\gamma_{\text{in}} = 0.0$ becomes obsolete as r_{s} shrinks and instead the outer $\gamma_{\text{out}} \sim 1.7$ (which is well constrained in AIC_{p}) is the important parameter that determines the actual slope behaviour of the halo mass distribution. An approximately identical profile could be generated by setting the

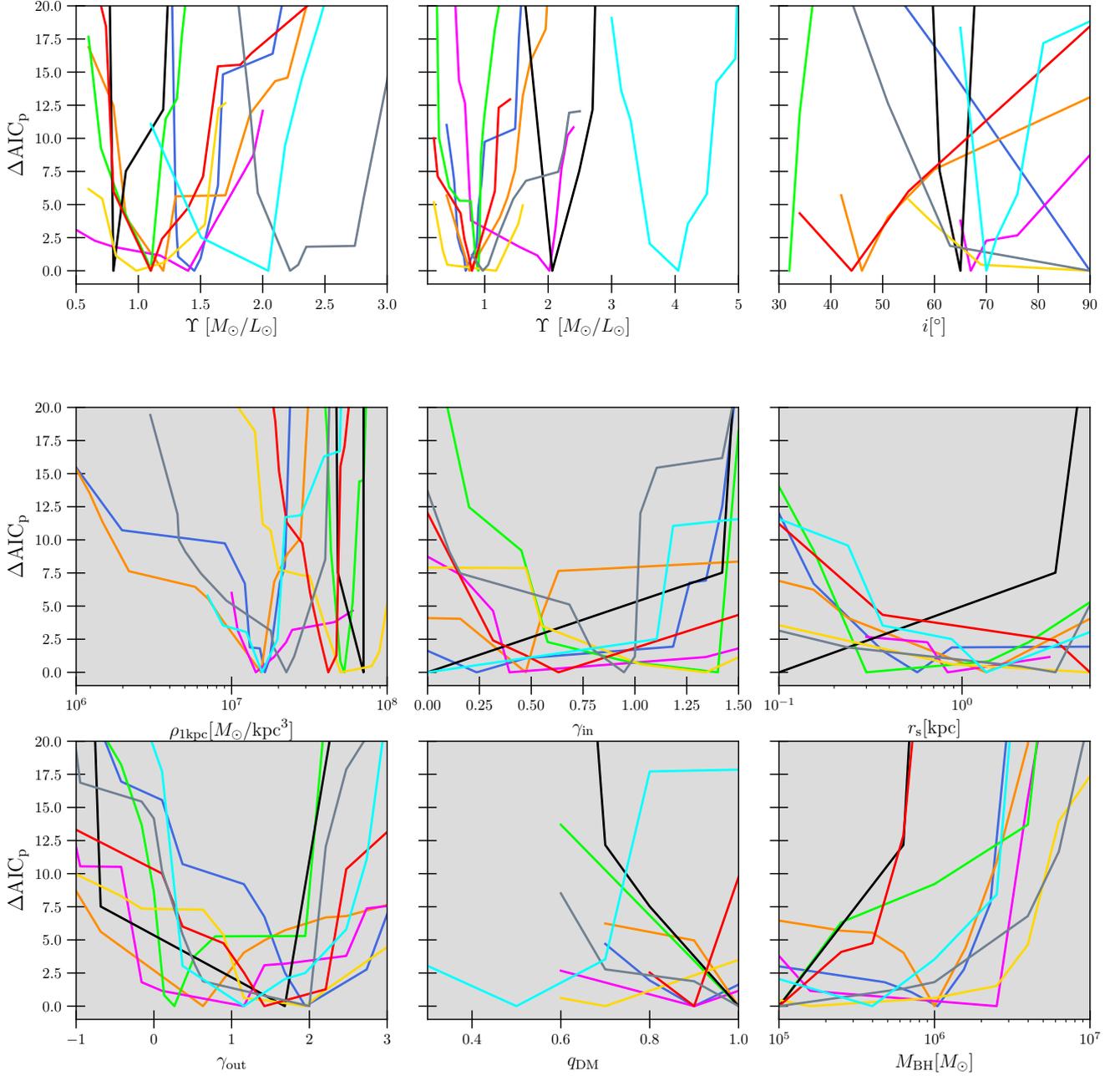


Figure 2. The ΔAIC_p envelopes of the orbit models probed for each dE. The upper 3 panels that are concerned with the luminous model components were discussed in VW-I. The gray-colored panels show the parameters that describe the ‘dark’ components of the orbits models, i.e. the black hole and dark matter. As argued in Sec. 3.2 the parameters that generate the halo distribution should be treated as nuisance parameters, specifically γ_{in} , γ_{out} , and r_s are very inter-correlated when the FoV is limited (as is always the case).

$\gamma_{\text{in}} \sim 1.7$ and moving the scale radius far outside, reaching the *upper* limits r_s , and consequently rendering γ_{out} irrelevant. Likewise, it is not very concerning that the outer asymptotic slopes γ_{out} appear to be implausibly shallow: the best models we found have $\gamma_{\text{out}} \in [0.2, 2.0]$, and that would mean the models have an infinite total mass when integrated out to $r \rightarrow \infty$. However, due to the degradation of the constraining power outside the FoV, mass models that differ only far outside the FoV are virtually indistinguishable for dynamical models. The γ_{out} constraints we measure instead describe, at best, the slope in the vicinity of the FoV edge, and even then only if the r_s is small enough for γ_{out} to be relevant. We will investigate the variability of the constraining power with radius on a methodological level in LT. To some degree this behaviour has been noted in the past several times, as it manifested as an overestimation of the total mass and uncertainty of the kinematic structure at large radii outside the FoV (e.g. Gerhard et al. 1998; Thomas et al. 2005).

4.2. Halo densities and dark matter fraction

Instead of interpreting these nuisance parameters we now focus on directly evaluating the mass distributions they generate. The *left* panels of Fig. 3 show the (spherically averaged) mass densities of the dark matter halo, the stars, and their combined density of the *best* dynamical models we found for each dE. In other words the Figure shows the actual mass distribution of the model with $\Delta\text{AIC}_p = 0$ in the nuisance parameter grid (Fig. 2). The stellar density shown here incorporates the spatially variable stellar mass-to-light ratios $\Upsilon_*(r)$ we equipped the models with, meaning it's not merely a deprojection of the photometry at some viewing angle. Fig. 4 displays the *total* enclosed mass $M_{\text{tot}} = M_{\text{DM}}(<r) + M_*(<r)$ vs radius in the left panel and in the right panel we quantify the relative contribution of the dark matter by displaying the cumulated dark matter fraction $f_{\text{DM}} = \frac{M_{\text{DM}}(<r)}{M_{\text{tot}}}$ as a function of radius. As a visual guide we also indicate the location of the stellar effective radius r_e in the mass profiles.

Despite the huge variety in candidate halo profiles we probed (see Fig. 1) the recovered density profiles ρ_{DM} show that the DM halos of the dEs behave qualitatively similar to each other, but, at the same time, are distinct from their baryonic counterparts. Dark matter is much less centrally concentrated and contributes only little to the total mass within the center, but its density profiles fall off less steeply. As such, the dark matter contribution to the total mass budget becomes increasingly more important with radius, as the dark matter fraction is surging outside radii greater than about $1r_e$.

In and of itself this is of course nothing new as numerous studies in the past decades have found this to be the case (starting from Rubin & Ford 1970). However, our results extend on the majority of previous studies in that we relax the assumption that the baryonic matter component follows the light distribution. Instead, we allow for a spatially variable *stellar* mass-to-light ratios (similar as Mehrgan et al. 2024 did for massive ETGs). This has the advantage that ‘missing’ mass, i.e. any difference between light and total dynamical mass is not necessarily counted as DM but that we can also account for changes in the stellar populations. The fact that we still see the self-similarity between the recovered halos and a systematic difference to the baryonic mass distributions is encouraging and increases the significance of the DM detection. If the stellar matter and dark matter components were completely *degenerate*, or if the missing mass was caused solely by a complex variability of the underlying stellar population, then we would expect a larger variability in the recovered halo masses, given the huge variety of candidate models that we probed (Fig. 1). For example, halos that increase in density with radius or halos that are congruent with the baryons (mass-follows-light) appear to be strongly ruled out by the dynamical modeling. All in all, the nuisance parameters that set up the halo may be fairly noisy (which is amplified by the large flexibility we allow in the mass models), but the need for an additional non-baryonic mass contribution that is distributed in a certain way is evident.

4.3. Density gradients - Dark matter vs Baryons

To quantify the degree to which the baryonic and dark matter differ we can evaluate their local density gradients as a function of radius. To this end we should focus on the density gradients in the regions where we believe the masses to be robustly constrained. As discussed in Sec. 4.3 the dynamical constraints are strongest where data coverage is good, and the dark matter fraction is high (see also LT). Therefore, to measure the slope in the most robust way, one could measure it within the aperture defined by the FoV and then weigh the local density gradient by the local f_{DM} . However, this turns out to be dangerous because the true f_{DM} itself is unknown and is a property that we estimate from the modeling itself. Therefore it relies on an accurate decomposition of baryonic and DM. Furthermore we may well expect that other galaxy types (e.g. ‘ordinary’ ETGs or dwarf spirals) have intrinsically very different dark matter fractions than dEs. Therefore if we would link f_{DM} and the definition of the slope we could find artificial correlations because both are not independent anymore.

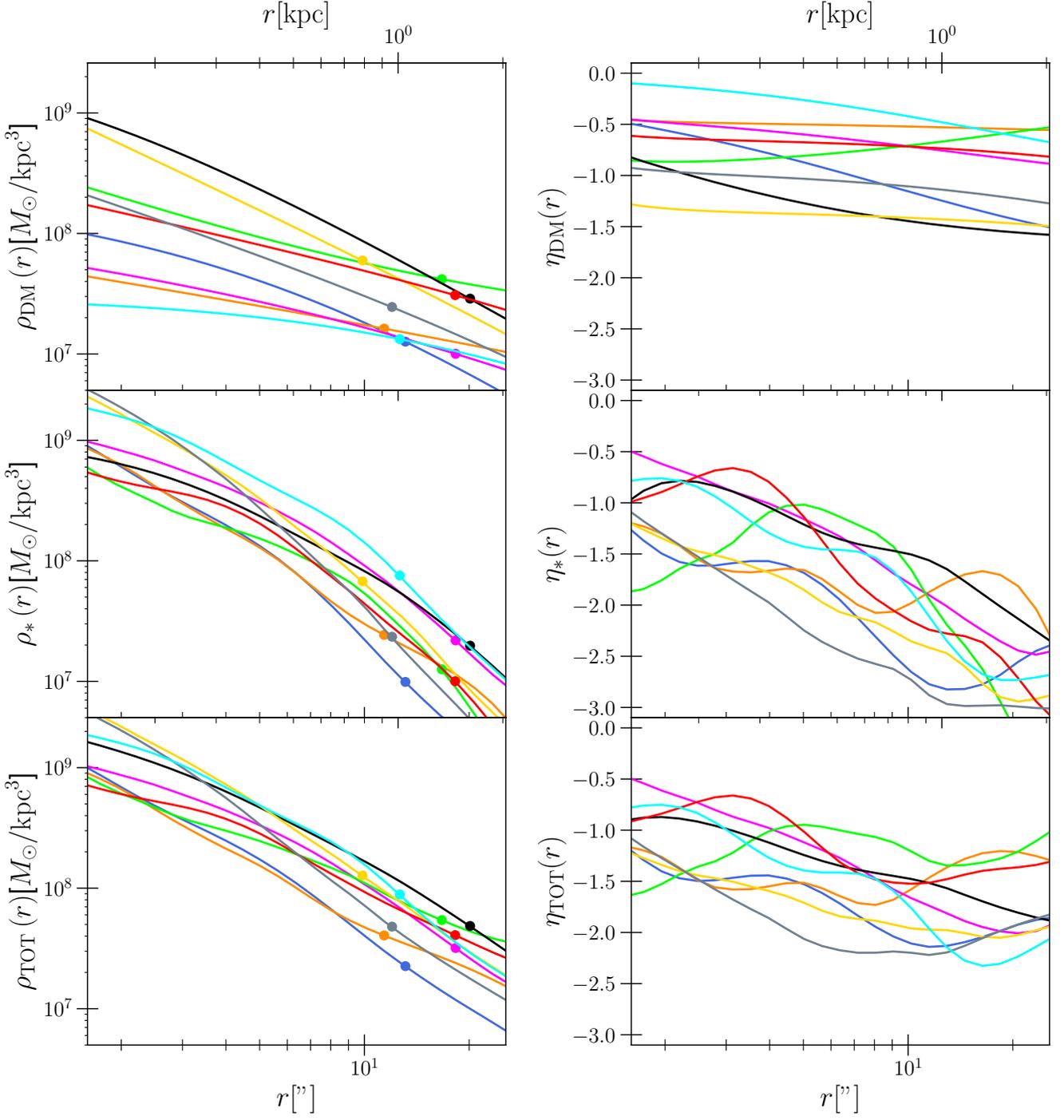


Figure 3. *Left side:* The (spherically averaged) matter densities ρ of each dE obtained from the best dynamical model we found. From top to bottom: dark matter density, stellar density (including the $\Upsilon_*(r)$), total mass density. We indicate the location of the (stellar) effective radius of each dE (cf. VW-I) by the point that overlap the corresponding curve. *Right side:* The corresponding (volume) averaged slopes η within spheres of radius r (see eq. 4) as a function of r . While the halo slopes stem from a parametric Zhao-description (i.e. the slopes are smooth by definition) the stellar component is non-parametric and is completely determined by the photometric deprojection, the inclination and the Υ_* -gradient of the best fit model. At a radius of $10''$, i.e. within the region where we expect the DM to be constrained best by the data (Sec. 3.3) the typical statistical 1σ -error of η_{DM} is $\sigma_\eta = 0.26$. The individual errors resolved for each dE can be inferred from Fig. 9 and Fig. 8.

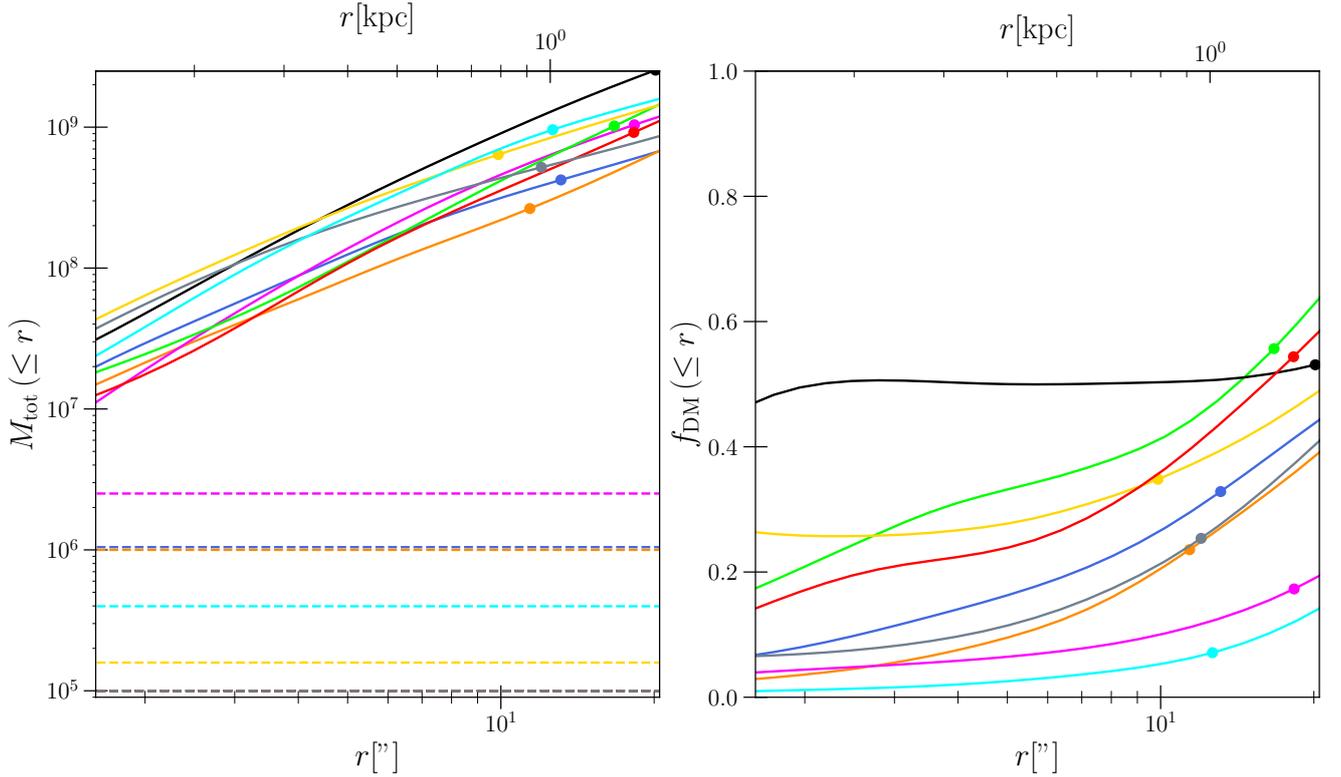


Figure 4. *Left panel:* The total enclosed mass of each dE vs radius (solid lines). *Dashed Lines:* The corresponding black hole masses of the best dynamical models. The spatial resolution of the VIRUS-W data is mostly limited by the fibre size and the seeing and is effectively of the order of $\sim 2''$, hence, we do not expect to resolve central black holes. While our models (formally) recovered black hole masses (Fig. 2) as shown here, their contribution to the total enclosed mass within the spatial resolution limit is negligible. Therefore we consider the best models to be essentially degenerate with any models with black hole masses $M_{\bullet} \lesssim 5 \cdot 10^6 M_{\odot}$. *Right panel:* The cumulated dark matter fraction within radius r . With some exceptions, the potential within the central few arcseconds of the dEs is dominated by the stellar mass.

To overcome these issues, we decided to calculate the *volume* averaged radial (logarithmic) slope within the sphere of radius r_{ev} :

$$\eta(< r_{\text{ev}}) = \frac{3}{r_{\text{ev}}^3} \cdot \int_0^{r_{\text{ev}}} dr r^2 \frac{\partial \ln(\rho)}{\partial \ln(r)} \quad (4)$$

where r_{ev} is the ‘evaluation’ radius. Several similar but not identical definitions and aperture conventions have been used in the past (e.g. Poci et al. 2017; Dutton & Treu 2014; Derkenne et al. 2023) to estimate the slopes of *total* matter densities. To place our results in a broader context and compare them to previous studies, we also analyzed the total slopes of our dE sample using some of the existing slope conventions (see Sec. 6.1).

We plot η as a function the $r = r_{\text{ev}}$ in the *right* panels of Fig. 3 for the dark matter and stellar components, as well as for the combined mass distribution. For the latter two the volume averaged slope is not an urgently needed measure since the constraints of total mass and baryonic mass do not scale with f_{DM} but we nonetheless display them here to enable a fair comparison between the different mass components³. The local density gradients of the two components show very different radial behavior. While both components have similar gradients in the galaxy centers, the curves diverge with increasing radius as the dark component barely gets steeper at larger radii. This is dynamically required as the models prefer total density gradients within the effective radius which are relatively shallow (notably shallower than an isothermal profile). However, the models achieve this without forcing the halo profiles to be extremely flat/cored ($\eta_{\text{DM}} \lesssim -0.5$).

4.4. Black holes - Are the centers dominated by luminous or ‘dark’ mass?

For the majority of the galaxies in our sample the halo contribution in the center is small to insignificant (Fig. 4) which suggests that the luminous baryons dominate the dynamics within the center. Still, the models formally also include a second ‘dark’ component, a central black hole that could affect the orbits of the stars tracing the center. The AIC_p constraints for the black-hole mass (Fig. 2) suggest a strong upper limit, with $M_{\bullet} \lesssim 10^6 M_{\odot}$. The weaker constraints towards lower black hole masses are within our expectations since the spatial resolution we achieve with the VIRUS-W data

would only allow us to resolve the sphere-of-influence of black holes with masses larger than $10^7 M_{\odot}$. This is illustrated in Fig. 4 which shows the enclosed mass of the total mass in comparison to the recovered black hole mass (horizontal lines). Adding or removing the recovered black hole masses to the total mass budget essentially has no impact on the total cumulated mass within our resolution limit ($\approx 2''$) that always exceeds $10^7 M_{\odot}$.

Our results suggest that neither DM nor black holes are a particularly relevant contribution to the potential in the center of most dEs (with some exceptions, e.g. VCC 1261). Instead the luminous matter distribution dominates the inner dynamics where the mass distribution is well approximated by the light without any need for additional ‘dark’ components (on the scales that we probed). Therefore we believe that the *upper* limits we obtained for black hole masses are relatively accurate and robust. All in all the results seem to suggest that over-massive black holes in dEs with $M_{\bullet} \gtrsim 5 \cdot 10^6 M_{\odot}$ are strongly ruled out by the dynamical modeling. In the context of the cusp-core problem (see below, Sec. 5.2) this already has implications since lower black hole masses are less likely to have affected the dark matter halos significantly, e.g. via Active galactic nuclei (AGN) feedback. Even though the fraction of dwarf galaxies that exhibit detectable activity is small in the local Universe, AGNs in dwarfs could have been more important at higher redshift (Mezcua et al. 2019; Sharma et al. 2022). Future studies with significantly higher resolution may even better constrain how important black holes are in dwarf galaxies. Our upper limits $M_{\bullet} \approx (10^5 - 10^6) M_{\odot}$ suggest that their influence on the DM distribution can not have been dramatic.

5. DARK MATTER IN QUIESCENT DWARFS - IN TENSION WITH STANDARD COSMOLOGY?

Apart from some notable exceptions (van Dokkum et al. 2018; Shen et al. 2021; Comerón et al. 2023) the vast majority of galaxies are known to require a dark matter halo (or a modification of gravity). In line with this, the dEs we investigated also exhibit a dynamical necessity for an additional dark component that increases the total mass budget (particularly at larger radii) beyond what the luminous mass distribution seems to suggest. However, with our advanced modeling technique, we attempt to go beyond merely demonstrating the dynamical necessity of dark matter in dEs: We try to accurately recover the exact profile, flattening and amount of dark matter in these galaxies. In principle, such distributions could even allow us to infer information on the underlying cosmology in which the

³This line of argumentation is of course also true for the stellar component, i.e. we expect the constraints on the stars to become worse where $f_{\text{DM}} \sim 1$. However, for our dE study this is not relevant since the baryonic contribution is sufficiently large throughout the entire FoV.

halos have formed. Within standard cosmology model (Λ CDM) structure formation on large scales is generally explained well but several problems and tensions at smaller scales were reported that seem to conflict with it (cf. Sec. 1), which will be investigated in the following.

5.1. Properties to analyze the DM structure

To address whether our dynamical constraints for the dEs are in tension with the standard Λ CDM paradigm, we first need to establish a set of properties that describe the halo succinctly. As discussed in Sec. 4 the DM halo is best constrained at radii where f_{DM} is large and within the FoV. For most of our sample this is roughly around $10''$ as the majority of dEs have negligible or low dark matter contribution in the center and our FoV covers radii out to $10''$ - $15''$ (cf. VW-I). Since we intend to compare the halos of each dE to one another and interpret their physical origin and evolution consistently we want to measure the average density and slope evaluated within a distance-independent, *identical* aperture for all galaxies. Therefore we decided to evaluate each halo within a circular aperture of the same intrinsic radius $r_{\text{ev}} = 0.8\text{kpc} \approx 10''$ which is at around one effective radius for the dEs.

Given such an aperture, we can then characterize the dark matter halos using the following **three quantities**:

i) The **average density** $\overline{\rho_{\text{DM}}}$ of the dark matter enclosed within the aperture:

$$\overline{\rho_{\text{DM}}} = \frac{3}{4\pi} \cdot \frac{M_{\text{DM}}(r < r_{\text{ev}})}{r_{\text{ev}}^3} \quad (5)$$

ii) The volume **averaged logarithmic slope** η_{DM} of the halo within the aperture, as given by equation eq. 4 evaluated at $r_{\text{ev}} = 0.8\text{kpc}$.

iii) The halo **flattening** within the aperture $r_{\text{ev}} = 0.8\text{kpc}$ which, in our case, is aperture independent and identical to the nuisance parameter q_{DM} because the models only allow for spatially constant axisymmetric dark matter flattenings.

5.2. The cusp-core problem in dEs

From Fig. 3 it is evident that the dark matter slopes of the dEs exhibit a similar diversity (sample scatter) as the slopes of their luminous stellar distributions do. But the dark matter distribution is (on average) shallower and changes less drastically with increasing radius than the stars. First and foremost, this result highlights the well known requirement of galaxies having a total mass that decreases less steeply with radius than the luminosity distribution would suggest (e.g. Rubin & Ford 1970). For many observed galaxies that are being studied with dynamical modeling this expresses itself in

the need for an almost entirely flat/cored halo model component that flattens the overall mass gradient. The resulting halo models are often times so cored that they are in tension with the Λ CDM simulations of structure formation which predict cuspier distributions, particularly in the regime of small dwarf galaxies and late-type galaxies (see Sec. 1).

However, in the regime of dEs the question whether the cusp-core problem persists, and if so how strong, is far from settled. While the dark matter in our dE models falls off less steeply with radius than the baryons (as expected), the halo density models are not completely flat either. There is no unequivocal preference of the halos being either very cored, or particularly cuspy. Instead, we find the slopes η_{DM} to be mostly *moderate* but with considerable scatter across the different dEs in our sample (Fig. 3). The most cored dEs have $\eta_{\text{DM}} \sim -0.5$ while the most cuspy ones have $\eta_{\text{DM}} \sim -1.4$. This sample scatter is larger than the typical 1σ -error $\sigma_{\text{DM}} = 0.26$ we find for the dark matter slopes and larger than the radial change in DM slope within one effective radius. Therefore we argue this sample scatter of the halo slopes is not a measurement uncertainty but displays the real diversity of the halos. In fact, we may expect exactly this level of diversity considering the corresponding level of variety in slopes of the stellar matter (middle panel of Fig. 3) which stems from the deprojection of the observed photometry (and indirectly its gradients). If the luminous matter exhibits this level of diversity, then it may be reasonable to expect that the halos of dEs can also exhibit a similar diversity in slopes.

Compared to the sample scatter across different dEs the radial change in slope is slightly smaller within the investigated radial range. Within one effective radius the change in slope is similar in scale to the 1σ error and, thus, barely statistically significant. Conversely, the slopes of the baryonic matter display a significant and systematic decrease over the entire range covered by kinematic data. Under the *assumption* that the true distribution of dark matter follows a profile with a *single* distinct scale radius, these results imply that the *scale radii* are dynamically unconstrained by our data. They must be located either far outside the FoV or in the very center where a negligible dark matter contribution (e.g. if $f_{\text{DM}} \sim 0$) could prevent a change in slope to be dynamically detectable by the models as it barely changes the net potential. The former case is more likely because the total enclosed mass would be infinite otherwise. While the detailed size of halos relative to their stellar half-light radius may depend on various factors (e.g. redshift, angular momentum, etc.), we can anticipate the halo size to be generally much larger than the stellar half-light

radius (e.g. Somerville et al. 2018). Consequently, we expect the scale radius to be far outside the $1r_e$ FoV, and we argue the nearly constant slopes we measure inside our FoV are mostly representative of the true *inner* logarithmic halo slopes. Strictly speaking, of course, this holds only if the true dark matter distribution has only a *single* scale radius. One could easily imagine a second or multiple slope transitions (i.e. scale radii). For example, a very small core within the regions where we measured, $f_{\text{DM}} \approx 0$ which we would not be able to detect dynamically with our current observations.

Given the above arguments, we interpret the dark matter density slopes we measured within the 0.8kpc aperture to be representative of the *central* slopes of the dark matter distributions. If we analyze the density slopes within this distance-independent aperture r_{ev} the sample average for the halo component is $\eta_{\text{DM}} = -0.91 \pm 0.35$, which is contrasted by a much steeper luminous matter slope with a sample average of $\eta_* = -2.04 \pm 0.40$. While we believe the slope within 0.8kpc is the most robustly constrained measure of the central dark matter slopes we can obtain from the dynamical models, the values are not very sensitive to the exact aperture that we use. For example, if we were to evaluate the dark matter slopes within $r_{\text{ev}} = 0.2\text{kpc}$ (i.e. close to the spatial resolution limit of our data) we would obtain an average dark matter slope of $\eta_{\text{DM}}(r = 0.2\text{kpc}) = -0.73 \pm 0.38$.

These slope estimates allow us to address the cusp-core problem from a new perspective, as detailed measurements of halos in this mass regime mostly stem from rotation curve modeling. This has a strong selection bias towards LTGs since it relies on the existence of a rotating HI disk. The dynamical models we employ here can complement these constraints, as they use stars as tracers of the gravitational potential instead. While stellar dynamical modeling has been employed to measure the halos of ETGs before, most studies are confined to more massive ETGs or smaller dwarf galaxies (dSphs) within the Local Group. Studies of early-types in the intermediate mass regime of the dEs are rare and, even then, do not attempt to explicitly measure the slope of the decomposed dark matter component but instead the *total* density slope (which we will also compare in Sec. 6.1). While a direct comparison to ETGs in the same mass range of our dEs galaxy is difficult, we can place our results in the context of the measured DM slopes of LTGs and the smaller resolved early-type dSphs of the Local Group.

While DM slopes for the most massive spirals are often ambiguous (some are cored other cuspy) the majority of studies of dwarf LTGs ($M_B \gtrsim -19\text{mag}$) tends to favor cored dark matter distributions (McGaugh & de

Blok 1998; Côté et al. 2000; de Blok & Bosma 2002; Gentile et al. 2004; de Blok et al. 2008; Donato et al. 2009; Plana et al. 2010; Oh et al. 2011b). For example, the dwarf irregulars (dIrr) analyzed by Oh et al. (2015), which are in the mass regime of our dEs, suggest a typical DM slope ~ -0.29 , and the dwarfs sample of Adams et al. (2014) has a mean DM slope of ~ -0.58 . In contrast, in the *quiescent* counterparts of the dIrr, the dSphs, the circumstances are much more ambiguous. First results of Walker & Peñarrubia (2011) using a chemo-dynamical approach seemed to have ruled out NFW-like halos for Fornax and Sculptor, though Genina et al. (2018) later showed that the models can possibly lead to mis-identification of cusps as cores. In a series of papers Jardel & Gebhardt (2012); Jardel et al. (2013); Jardel & Gebhardt (2013) have modelled dSphs of the Milky Way (Carina, Draco, Fornax, Sculptor, and Sextans) using Schwarzschild models. They found a diverse range of central DM profiles, with some being cuspy while others being cored. They concluded that (at least on average) the central halo slope scatters around an NFW-like profile. Similarly, Hayashi et al. (2020) found that the halos of 8 dSphs show a significant variety but tend to favor more cuspy density distributions. Recently, De Leo et al. (2023) analyzed resolved stellar motions in the Small Magellanic Cloud using Jeans modeling and found a cuspy structure that is more consistent with theoretical predictions.

Taking into account the results of our dEs and the ambiguity in the dSphs samples, it appears the results for ETGs samples are not as suggestive of a core-cusp problem as the observational constraints for LTGs are. This disparity between LTG and ETG samples could be an artifact from *systematic* differences between the modeling methods that are being employed. Many studies test parametric models such as NFW-profiles, or non-singular isothermal/logarithmic profiles, which may or may not be flexible enough to capture the structure of real DM halos. The effects of halo parametrization have not yet been investigated systematically, and it could be the case that some cored halo models nominally provide a better fit but underestimate the central DM density considerably. In LT we will investigate the effects of halo parametrization on a methodical level. Apart from this parametrization issue that likely extends to all dynamical modeling techniques, the observations of cores could also be facilitated by the fact that the majority of LTG studies are conducted using gas rotation curves whereas ETGs are constrained using stars as tracers. However, for a sample of *late-type* dwarfs Adams et al. (2014) investigated whether dynamical models that use gas as tracer for the gravitational potential recover different

DM slopes than Jeans models (with stars as tracers). Using generalized NFW (gNFW) models they find both approaches to be fairly consistent with the gas models having a sample averaged central slope $\gamma = -0.67 \pm 0.1$ and the stellar models $\gamma = -0.58 \pm 0.24$. This suggests that the apparent difference in dark matter slopes of early- and late-types does not stem from an intrinsic bias/difference between the stellar vs gas dynamical models. Still, further comparisons of different modeling techniques applied to the same galaxy are required to judge about modeling systematics.

If for now we assume systematics are negligible and results for LTGs and ETGs can be trusted equally, this slope-morphology dichotomy may well be genuine because the two classes can be expected to follow very different evolutionary paths. The correspondingly different baryonic feedback and/or environmental effects could have driven initially identical dark matter halos to diverge over time. In other words, even if all DM halos had initially NFW-like slopes, we may expect that LTGs observed today are on average more cored than the dEs. Since we argue the amount of diversity in measured slopes within our dE sample is not driven by noise or systematic errors but reflects the true diversity that dEs exhibit (see above), the even larger slope difference to the LTG studies strongly suggests that there must be mechanisms (ab initio or secular ones) that discriminate between different galaxies, leading to some halos being shallower than those of others (see also Sec. 6).

The situation whether dark matter in quiescent dwarfs is, in fact, cored enough to be considered in tension with standard cosmology appears to be much more ambiguous than it is for the LTGs. A considerable amount of ETGs studies suggest mildly cored and near NFW slopes. Still, the typical central slopes that we measured $\in [-0.7, -0.9]$ are in mild to moderate tension with the vast majority of DM-only simulations (DMO) of Λ CDM structure formation. Early DMO studies predicted (often universal) cuspy halo profiles with central slopes ranging from -1 to -1.5 (Navarro et al. 1996a, 1997; Moore et al. 1998; Fukushige & Makino 2001), and even more recent DMO simulations with higher resolution struggle to reach slopes much flatter than -0.8 (Gao et al. 2008; Stadel et al. 2009). While the tension between the DM slopes of simulations and observations is significantly mellowed in the quiescent dwarfs, the existence of galaxies with DM slopes of -0.5 nonetheless rejects DMO simulations, suggesting at least some requirement for modification. Hydrodynamic simulations which include the response of the DM to baryon physics and friction appropriately may (possibly more comfortably than for LTGs) explain these remaining differences

(e.g. El-Zant et al. 2001; Mashchenko et al. 2006; Del Popolo 2009; Governato et al. 2010; Cole et al. 2011; Governato et al. 2012; Nipoti & Binney 2015; Orkney et al. 2021).

All in all, neither the simulation nor the observation side seem to be entirely decided whether the halos of dwarf galaxies are cuspy or cored. In the last decades several solutions have been proposed to reconcile the observations with simulations (for a review see Del Popolo & Le Delliou 2021). Broadly speaking, solutions can be categorized as follows: i) Baryonic feedback/outflows which transfer to the DM. For example, star formation bursts could have removed parts of the baryons, and in the course of that also some of the dark matter, leading to the observed cores (e.g. Navarro et al. 1996a; Gnedin & Zhao 2002; Governato et al. 2010; de Souza et al. 2011; Oh et al. 2011a; Madau et al. 2014). Similarly, interactions with the environment such as mergers, harassment or ram-pressure stripping (directly and indirectly by regulating star formation) could have affected the dark matter profiles (e.g. Del Popolo 2012). ii) Cosmological solutions: Theoretical concepts such as self-interacting dark matter, fuzzy dark matter, superfluid dark matter, could naturally produce cored halo profiles (e.g. Spergel & Steinhardt 2000; Harko 2011; Robles & Matos 2012; Elbert et al. 2015) iii) Modifications of Newtonian gravitation can make cuspy halos appear to be cored in a Newtonian analysis (e.g. Benetti et al. 2023).

Overall we conclude that the cusp-core tension for dEs is not as severe and standard Λ CDM simulations that accurately model the effects of baryons may well be able to explain the halo distributions without the need for invoking more exotic physics. In Sec. 6 we will explore (under the umbrella of standard Λ CDM cosmology) whether our results point towards a specific baryonic mechanism that is driving the diversity in measured DM slopes of the dEs and the systematic difference to LTGs.

5.3. Are DM halos spherical?

Numerical Simulations of DM halo formation not just predict their central slopes but also their 3D shape. Therefore the axisymmetric flattening q_{DM} we measured for the dEs (Fig. 2) puts us in the position to probe cosmological predictions and halo assembly further. The flattening of dark matter halos (other than the Milky Way's halo) is rarely probed by dynamical modeling studies. In App. A we demonstrate that dark matter flattening is detectable using our dynamical modeling setup and, in fact, plays a significant role in determining the overall quality of the model. Even though the N-body simulation in App. A is slightly triaxial and

changes shape with radius the constant, axisymmetric flattening $q_{\text{DM}} = 0.8$ we recovered approximates the average shape of the halo (to first order) well and unbiased.

For dEs we find that dark matter and stars are not only different with regard to their *radial* distributions, but they also have systematically different flattenings. Fig. 5 compares the average *intrinsic* axis ratios of the luminous matter q_* with that of the dark matter q_{DM} . For all galaxies, but VCC 1528, the halo appears to be rounder than the stellar component it hosts. In fact, the majority of the DM halos are spherical or close to it. The only halo that exhibits a strong amount of halo flattening is that of VCC 2048. However, VCC 2048 also has the flattest stellar distribution ($q_* \simeq 0.4$) and the lowest dark matter fraction with $f_{\text{DM}} < 0.1$ throughout the entire investigated radial range (cf. Fig. 4), hence, the central halo shape may not be very well constrained.

DMO simulations usually predict complex halo shapes that can be prolate or triaxial with significant shape variation at different radii (e.g. Frenk et al. 1988; Allgood et al. 2006; Hayashi et al. 2007), therefore it may be surprising that almost the entirety of the dE sample is fitted best with nearly spherical dark matter halos. However, these DMO predictions are also in tension with most observational results for the Milky Way’s DM halo (e.g. Ibata et al. 2001; Law & Majewski 2010; Vera-Ciro & Helmi 2013; Bovy et al. 2016; Wegg et al. 2019). These suggest a nearly spherical, oblate dark matter structure despite the Milky Way’s much flatter stellar structure. For example, by analyzing the *Gaia* proper motions of RR Lyrae stars Wegg et al. (2019) have found an ellipsoidal flattening $q_{\text{DM}} = 1.0 \pm 0.09$ of the Milky Way halo out to radii of 20kpc.

Similar to the cusp-core problem, this tension with the DMO simulations is significantly mellowed when baryonic interactions are included in the numerical simulations: the halo shapes are transformed by the interactions with baryons to be more spherical/oblate (e.g. Katz & Gunn 1991; Chisari et al. 2017; Chua et al. 2019; Chua et al. 2022; Cataldi et al. 2023; Orkney et al. 2023). This reflects that DM particles that reach the center on box orbits are gravitationally affected by the condensing baryons such that their orbits change and become more tube-like and circular (cf. Debattista et al. 2008). This modification of the flattening caused by the baryons is most dominant in the center of the galaxies: while DMOs predict prolate distributions in the center which only become more spherical at large radii the baryon inclusion make the center more spherical which overall leads to a more radially constant and close-to-spherical halo distribution (Abadi et al. 2010). Hydrodynamical simulations not only predict that the result-

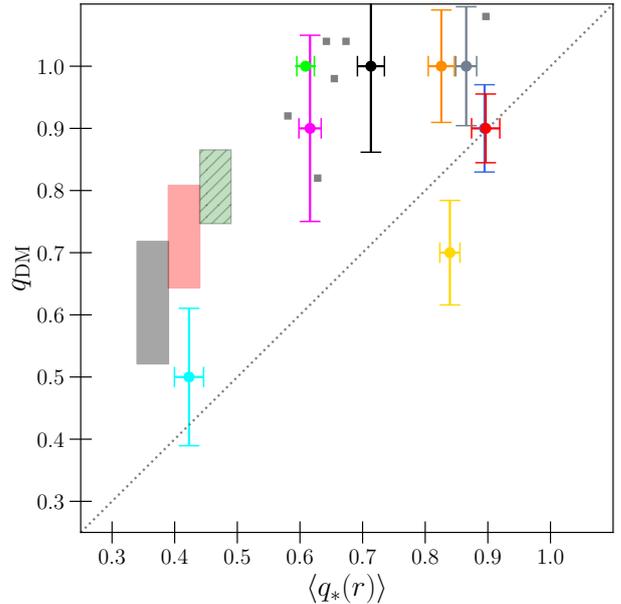


Figure 5. The *intrinsic* axisymmetric flattening of the DM halo q_{DM} vs the corresponding (radially averaged) q_* of the stars for the dEs. Also included are the dynamically recovered axis ratios of the dSph of the Milky Way (gray squares) from Hayashi et al. (2020). We indicate the *expected* q_{DM} from the Illustris simulations (Chua et al. 2019) for their DMO runs (gray bar), for the full hydrodynamical run in the dE regime $M_* \sim 10^9 M_\odot$ (red bar) and for higher masses ($10^{11} M_\odot$) at which their halos are roundest (green bar). They describe the halo triaxial by minor/intermediate axis-ratios q and s , to compare it to our axisymmetric values, the mean of q and s is shown here. The bar height represents the 25th-75th percentiles of their galaxies. The placement of the bars on the q_* -axis is arbitrary because they do not resolve q_* , but from the analysis of more massive ETGs in Illustris TNG100 by Pulsoni et al. (2021) we may expect $q_{\text{DM}} \geq q_*$ with a large range of stellar flattening $0.2 \leq q_* \leq 0.9$. Congruent with these simulations, our dEs have $q_{\text{DM}} \geq q_*$, but have even rounder halos than statistically expected.

ing halo becomes more spherical, but also that they end up to be rounder (by about $\Delta q \sim 0.2$) than the baryonic distribution they host (e.g. Tenneti et al. 2014; Pulsoni et al. 2021). This is congruent to our finding that the dark matter is more spherically distributed than the baryonic mass (Fig. 5).

In Fig. 5 we indicate the range of halo flattenings we may expect from numerical Λ CDM simulations by showing the median flattening of galaxies in the Illustris simulation analyzed by Chua et al. (2019). The inclusion of baryons typically changes the axis ratios of the simulated galaxies by about +0.1 to +0.2 which makes the average halo to have $q_{\text{DM}} \approx 0.75$. Like it is the case for the Milky Way’s halo, the inclusion of baryons does lift some of the tension, but our observational constraints

still suggest surprisingly spherical dark matter distributions. 7 out of 9 dEs have a q_{DM} larger than the 25-75th percentile of the Illustris simulated galaxies. The analysis of Chua et al. (2022) suggests that the effect of baryon-induced dark matter flattening becomes smaller as the total halo mass is decreased and/or as baryonic feedback parameters, such as stellar wind or black hole feedback strength, are increased. The former implies that we may expect galaxies in more massive halos to be even more spherical, while the latter implies that galaxies with stronger baryonic feedback are more prolate as they decrease the stellar mass fraction. Still, the observational constraints of our dE sample suggest that even halos in quiescent dwarfs are similar to the Milky Way in that they follow a nearly spherical distribution. Whether one can detect a dependency of the sphericity on the total mass or morphology (like it seems to be the case with the density slopes) remains to be seen at this point, as observational constraints need to be extended to more galaxies.

The fact that some statistical tension between hydrodynamic simulations and our observation persists could be explained by the approximations and assumptions that were made to obtain the observational constraints. Firstly, our small sample of dEs may not be overly representative for the average galaxy in this mass regime. The majority of our sample is observed to have a fairly round light distribution already (cf. ellipticities in VW-I) which could also favor more spherical DM halos, i.e. q_{DM} could be selection biased. Furthermore, considering baryons are a driving factor in determining the halo shape, we may also expect there to be significant differences between different galaxy morphologies⁴. Secondly, the dynamical models assume an axisymmetric shape with a radially constant flattening, while real halos are probably triaxial with radially changing shape. This could bias our results for q_{DM} to be more spherical, though the test on the N-body simulation (App. A) does not suggest so. Thirdly, especially for low mass halos, the baryon distribution and its embedding dark matter halo may be misaligned significantly (Chisari et al. 2017), an effect which our dynamical models can not emulate at this point.

If, on the other hand, the observational constraints are robust and the hydrodynamical simulations accurately model baryonic interactions, the measured shapes of DM halos could be a strong indication of a deviation from

Λ CDM cosmology. The excess in sphericity could be a direct consequence of non-CDM particles. For example, self interactions of the dark matter particles can heat up their orbits as they pass through the dense halo center and interact with each other (e.g. Peter et al. 2013; Vogelsberger et al. 2016; Brinckmann et al. 2018). As a result, the halos become rounder in the center, which is where we constrain the dE halos. Similarly to self-interacting dark matter, Fuzzy Dark matter particles can also produce more spherical halo shapes (Marsh & Silk 2014; Dutta Chowdhury et al. 2023).

At this point in time, observational attempts at measuring the intrinsic shape of DM halos are still in the beginning stages. The few constraints that exist are for nearby spirals using *edge-on* rotation curve fitting (e.g. Peters et al. 2017), for Local Group dSphs using Jeans modeling (e.g. Hayashi et al. 2020), or are for the Milky Way’s halo inferred from globular cluster and stellar streams (e.g. Ibata et al. 2001; Law & Majewski 2010; Vera-Ciro & Helmi 2013; Bovy et al. 2016; Posti & Helmi 2019; Wegg et al. 2019). While gravitational lensing can provide independent shape estimates at larger distances, it can only directly constrain the *projected* flattening of the mass distribution. Constraining the halo shapes with stellar dynamical modeling could therefore be an invaluable additional probe for early-type galaxies and further our understanding of cosmology.

Our dynamical constraints corroborate the observational results found for the Milky Way’s halo in that the dark matter is distributed close to spherical even though the stars occupy orbits in a more flattened distribution. However, as outlined above, several simulations and theories may explain such observations. To draw definite physical conclusions on the implications of the measured dE halo shapes, larger sample sizes with FoVs beyond $1r_e$ and dynamical models that are triaxial and/or allow radially changing shapes are required. Similarly, galaxy samples at different mass and redshift scales could also be helpful in our understanding. We only probe dEs with $M_* \sim 10^9 M_\odot$ in this study, but the ability/efficiency, e.g. of baryonic feedback, to make the halos more spherical is expected to vary with total mass, stellar mass fraction and redshift (e.g. Allgood et al. 2006; Chua et al. 2019; Chua et al. 2022). Investigating these features of the halo shape may allow us to differentiate between the different cosmological scenarios in the future.

6. HALO STRUCTURE - AN IMPRINT OF GALAXY FORMATION AND EVOLUTION?

The discussion in Sec. 5 has shown that the Λ CDM cosmology is not strongly ruled out by the observational

⁴ E.g. between ETGs and LTGs of the same mass. Though we may expect from the comparison of central slopes (Sec. 5.2) that q_{DM} of late-type dwarfs is even more spherical than early-type dwarfs.

shape and slope constraints we found for the dark matter halos of dEs. While there remains some tension (particularly in the halo flattening) numerical simulations that accurately model baryonic and environmental effects may well be able to explain the observations without the need for invoking more exotic physics. In the following few sections, we will explore what the recovered dark matter distributions can tell us about the evolution and formation of dEs, given we assume Λ CDM is indeed an accurate description of underlying cosmology. We will preface this analysis by a discussion of the total densities (stars and DM combined) which complements the discussion of the individual DM halos and facilitates comparison with the existing literature.

6.1. The total density slopes of dEs

The DM density slopes we reported in Sec. 5 rely on a robust decomposition of the DM halo from the stars and stellar remnants. While we stress-tested this on a simulation (App. A) there is of course no guarantee that this always works for real galaxies. However, under the assumption the Λ CDM paradigm is correct, we can also investigate galaxy evolution and differences between different morphologies without the explicit requirement for an accurate decomposition if we analyze the *total* mass distribution. The total mass is more strongly constrained by the dynamical models than the individual mass components are (see LT) and its recovery is not reliant on a successful decomposition. The same will hold true for the corresponding *total* density slopes. Furthermore, compared to the analysis of decomposed DM, a manifold of published studies exist that investigate the total densities but not the dark matter component on its own. In the following, we complement our analysis of the DM slopes (Sec. 5.2) by comparing the total density slopes of our dEs to those of published literature. This has the goal to place the dEs in a broader context, compare them to different galaxy classes and investigate possible formation mechanisms. Various conventions of measuring the total slopes (calculated within different apertures) are used in the literature. To make the comparison straightforward, we calculate the slopes for our dEs as they are defined in the studies that we compare them to in the following section.

6.1.1. Comparison with ‘ordinary’ ETGs

In Fig. 6 we display γ_{tot} , the mean logarithmic density slope of the total mass within the effective radius (see Poci et al. 2017), vs the effective velocity dispersions σ_e and the dark matter fraction f_{DM} for our dE sample and studies of ‘ordinary’ ETGs ($M_* \gtrsim 10^{10} M_\odot$). For the latter, we show three different samples, each based on a different modeling technique: i) Lensing models, ii)

Schwarzschild models, and Jeans anisotropic modeling (Cappellari 2008). Independently of the applied modeling method, the density slopes of the ‘ordinary’ ETGs congregate at $\gamma_{\text{tot}} \approx -2.1$ on average. In stark contrast, the slopes of our dEs are noticeably shallower (with a sample average of $\gamma_{\text{tot}} = -1.51 \pm 0.24$) and concentrate outside the scatter of ‘ordinary’ ETGs.

In the high mass range $\sim 10^{10} - 10^{12} M_\odot$ (i.e. $\log_{10}(\sigma_e) \gtrsim 2$) it is well established that (with little scatter) galaxies have around isothermal total density profiles at one effective radius. Because the steeper baryonic and shallower dark matter components seemingly ‘know’ of each other so as to always produce a nearly isothermal density together, this is also known as the ‘bulge–halo conspiracy’ (e.g. Gerhard et al. 2001; Thomas et al. 2007b; Koopmans et al. 2009; Auger et al. 2010; Dutton & Treu 2014). However, this ‘bulge–halo conspiracy’ actually only seems to hold in a limited mass regime. Several studies (Poci et al. 2017; Tortora et al. 2019; Li et al. 2019) have noticed that below $\log_{10}(\sigma_e) \sim 2.1$ the total slope and velocity dispersion (or stellar mass) of a galaxy are anti-correlated (i.e. less massive ETGs are shallower). As the total mass is increased, the galaxy profiles become steeper until near $\log_{10}(\sigma_e) \sim 2.1$ the typical slope reaches a minimum at $\gamma_{\text{tot}} \approx -2.3$, at which point the $\gamma_{\text{tot}}-\sigma_e$ breaks and flattens or arguably even reverts as galaxies with even higher masses become slightly more shallow again.

This break point near $\log_{10}(\sigma_e) = 2.1$ and $\gamma_{\text{tot}} = -2.3$ can also be identified in Fig. 6 as more massive galaxies do not become cuspier. If, on the other hand, the $\gamma_{\text{tot}}-\sigma_e$ correlation of the ETGs with $\log_{10}(\sigma_e) \leq 2.1$ is extended to the lower dispersions of our dE sample, then, the dEs are arguably exactly where one would expect them to be. This may suggest that the dEs, together with slightly more massive ETGs ($\log_{10}(\sigma_e) \lesssim 2.1$), form a continuous sequence for which the total slope γ_{tot} systematically decreases with increasing velocity dispersion σ_e . Whether continuously connected or not, the dEs have much shallower density distributions than ‘ordinary’ ETGs.

Conversely to the ‘ordinary’ ETGs, the total slopes of dEs do not fall on the same correlation of the dark matter fraction f_{DM} with the density slope (right panel of Fig. 6) as their shallower slopes would require a much higher dark matter fraction around $f_{\text{DM}} \sim 0.8$ than what is measured. In other words, the shallow γ_{tot} of dEs is not *solely* a result of an increased contribution of the halo to the total density within the effective radius. The strong correlation of total slope and DM fraction in ‘ordinary’ ETGs may be expected. Simulations suggest both γ_{tot} and f_{DM} increase with stellar mass (e.g. Lovell et al. 2018; Mukherjee et al. 2022) for galaxies

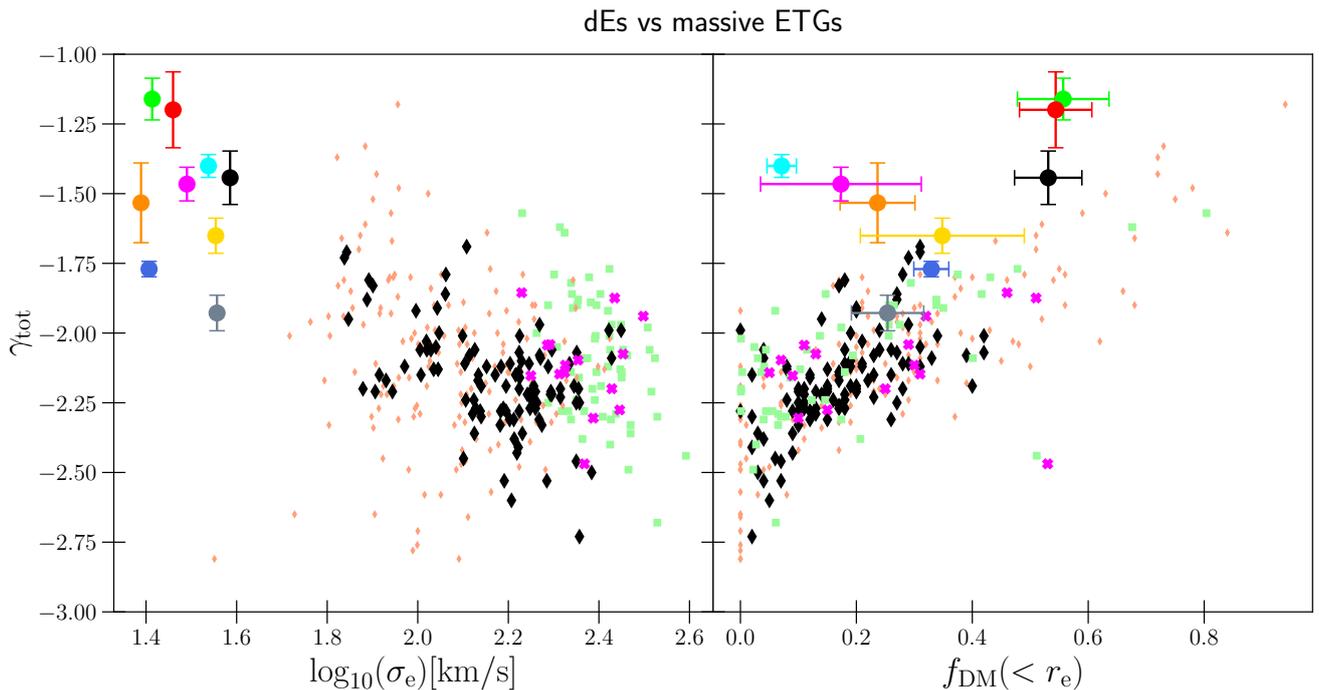


Figure 6. *Left panel:* The average logarithmic power-law slope γ_{tot} of the *total* mass distribution evaluated between $0.1r_e$ and $1r_e$ vs the luminosity-weighted velocity dispersion σ_e within one effective radius. *Colored Dots:* Our dE sample. *Diamonds:* The massive ETGs, part of the ATLAS^{3D}-survey (Cappellari et al. 2011), for which we obtained the slopes of model 1 in Poci et al. (2017). Their aperture varies slightly for each galaxy since it depends on the FoV coverage. But for the majority their γ_{tot} is essentially measured between $0.1r_e$ and $1r_e$. Following Poci et al. (2017) we differentiate between ATLAS^{3D} galaxies with good data quality (*Black diamonds*) and poor quality (*Red diamonds*). This data quality for each ETG was evaluated in Cappellari et al. (2013a). *Green squares:* The power-law slopes obtained from Auger et al. (2010) for the strong lensing ETGs which are part of the SLACS survey (Bolton et al. 2008; Auger et al. 2009). *Magenta crosses:* The total slopes calculated from the mass profiles of Thomas et al. (2007b) which were obtained with Schwarzschild which are based on an earlier version of the axisymmetric orbit code we use in this study. *Right panel:* The total slope γ_{tot} vs the dark matter fraction f_{DM} enclosed within $1r_e$ for the same data that was shown in the left panel. For the SLACS survey we obtained dark matter fraction from Posacki et al. (2014). For more published data of giant ETGs and comparison with cosmological simulations see also Fig.11 of Derkenne et al. (2023).

with masses above $\gtrsim 3 \cdot 10^{10} M_{\odot}$. If f_{DM} and γ_{tot} both correlate with stellar mass then they will also correlate with each other as seen here. We conclude the total density slopes of dEs arguably lie on the extension of ‘ordinary’ ETGs to lower masses, but their slopes and dark matter fractions are not as strongly correlated as those galaxies.

6.1.2. Comparison with quiescent and star-forming dwarfs

The above anti-correlation of total density slope and stellar dispersion (or stellar mass) has also been noted in LTGs, albeit with some offset/modification (Li et al. 2019; Tortora et al. 2019). In Fig. 7 we compare our results to those of Tortora et al. (2019). They analyzed and compared the total density slopes of 3 different galaxy samples: Another Virgo dE sample (Toloba et al. 2014; see also VW-I for a comparison), a LTG sample which includes star-forming galaxies ranging from S0

to Im (Lelli et al. 2016), and ‘ordinary’ ETGs (La Barbera et al. 2010; Tortora et al. 2012). They measured the total density slope (within the one effective radius) using the mass-weighted density slope γ_{MW} (cf. Koopmans et al. 2009; Dutton & Treu 2014; Tortora et al. 2014) which is defined as:

$$\gamma_{\text{MW}} = \frac{1}{M(r)} \int_0^r -\gamma(x) \rho(x) 4\pi x^2 dx \quad (6)$$

where $M(r)$ is the total cumulated mass, ρ the density and γ the corresponding logarithmic slope. For a single

power-law profile γ_{MW} equals γ_{tot} , but generally $\gamma_{\text{MW}} \geq \gamma_{\text{tot}}$ holds true⁵.

Similar to ETGs (Fig. 6) the star-forming galaxies also become cuspier as the total stellar mass (or dispersion) is increased but, in contrast to the ETGs, they exhibit no clear correlation-break at higher masses. The total slopes and dark matter fractions we recovered for our dEs are in remarkable agreement with the dE sample analyzed by Tortora et al. (2019). Consequently, dEs of both samples are systematically offset compared to LTGs of the same stellar mass, with dEs having steeper slopes by about ~ -0.5 on average. The total slopes we measure are of course spherically averaged quantities. While we may expect both, ETGs and LTGs, to sit in relatively spherical DM halos (Sec. 5.3) at least the stellar density of LTGs can be expected to be flatter than that of dEs. Still, if we assume the worst case where both stars *and* DM are distributed in an exponential disk profile, we expect their intrinsic circular speed curves $v_c(r) = \sqrt{GM/r}$ to always stay within 15% of the purely spherical equivalent (cf. Binney & Tremaine 2008; Tortora et al. 2019). Compared to the large γ_{MW} difference between LTGs and ETG this effect is miniscule.

dEs are also offset to LTGs in their dark matter fractions. Most dEs have a dark matter fraction around 30 – 40%. While the LTGs explore the full range of f_{DM} their distribution is lopsided towards higher dark matter fractions. Particularly LTGs in the mass regime of the dEs have typical dark matter fractions of around 70 – 80% (see also Sharma et al. 2023a). In contrast to the LTGs, the two dE samples occupy a narrow range in $\gamma_{\text{MW}} - f_{\text{DM}}$ -space with both properties slightly correlated.

Considering DM halos are generally thought to be shallower than the stellar distribution (e.g. Sec. 4.3) one could argue that the shallower *total* slopes of LTGs compared to dEs are a result of their higher dark matter fraction, whereas stellar differences in two morphologies play a subordinate role. However, it may not be as straightforward to draw conclusions from the total slopes alone. Both, f_{DM} and γ_{MW} , are measured within one stellar effective radius and are therefore aperture dependent, i.e. they depend on the concentration of the stellar distribution. However, at the same stellar mass the effective radii of LTGs may well be systematically

larger than those of the dEs (Tortora et al. 2019; Li et al. 2019), which may impede a consistent f_{DM} and γ_{MW} comparison of dEs and LTGs at the same *stellar* mass if the dark matter halo in LTGs is not similarly extended.

6.1.3. *The dichotomy in total slopes - A product of baryonic in- and out-flows?*

As discussed in Sec. 6.1.1 the total slopes of our dEs are not in tension with the existing evidence (Poci et al. 2017; Tortora et al. 2019; Li et al. 2019) for a *dichotomy* in the early-type sequence below/above $M_* \approx 10^{10} M_{\odot}$. It appears that below this mass threshold the slopes become continuously shallower at least until $M_* \approx 10^{8.5} M_{\odot}$, whether and how this changes for even lower mass remains open for now. The breaking point of this dichotomy in the ETG sequence has long been known to be a crucial point of change in many other galaxy properties as well (e.g. Bender 1988; Bender et al. 1989; Kormendy & Bender 1996; Kormendy 1999; Dekel & Birnboim 2006; Tortora et al. 2010; Kormendy & Bender 2013; Cappellari et al. 2013b; Nelson et al. 2018).

This dichotomy in the total slopes of early-type sequence and the offset (in slope and f_{DM}) of dEs compared to late-type dwarfs can be interpreted and explained by a coherent (but not necessarily comprehensive) picture of galaxy formation and processing. Depending on the galaxy’s total (stellar) mass, several processes could lose or gain in significance such as to produce the observed dichotomy in density distributions. In the following we briefly summarize important effects for different masses with an emphasis on the low-mass regime occupied by the dEs:

i) The *high-mass regime* $M \gtrsim 5 \cdot 10^{10} M_{\odot}$ at the turnover/breaking point the ETGs are slightly sub-isothermal ($\gamma \sim -2.3$). ETGs with higher mass are increasingly a product of (multiple) *dry* mergers, and as these dry merger stack up, the slope of the merger product gets more shallow and close to isothermal with $\gamma \sim -2.0$ (Remus et al. 2013, 2017). This reverts the correlation of total slope and stellar mass and breaks the correlation that is found for lower masses. In parallel to this merger effect, the feedback of AGNs could play an additional driving factor in making the giant ETGs shallower than ETGs at the breaking point. This is because AGN feedback becomes more efficient in suppressing star formation at higher masses (Tortora et al. 2010).

ii) In the *intermediate mass-regime*, near and around the turn-over point, the situation is likely complex as different mechanism are superimposed and possibly counter-acting each other. Nonetheless, we want to

⁵ The mass-weighted slope is essentially the analog of the volume averaged slope we introduced in eq. 4 but weighted by the local density. We argue for the dark matter analyzed on its own the volume averaged slope η is the safer choice (cf. Sec. 4.3), while for the total density γ_{MW} is more suitable.

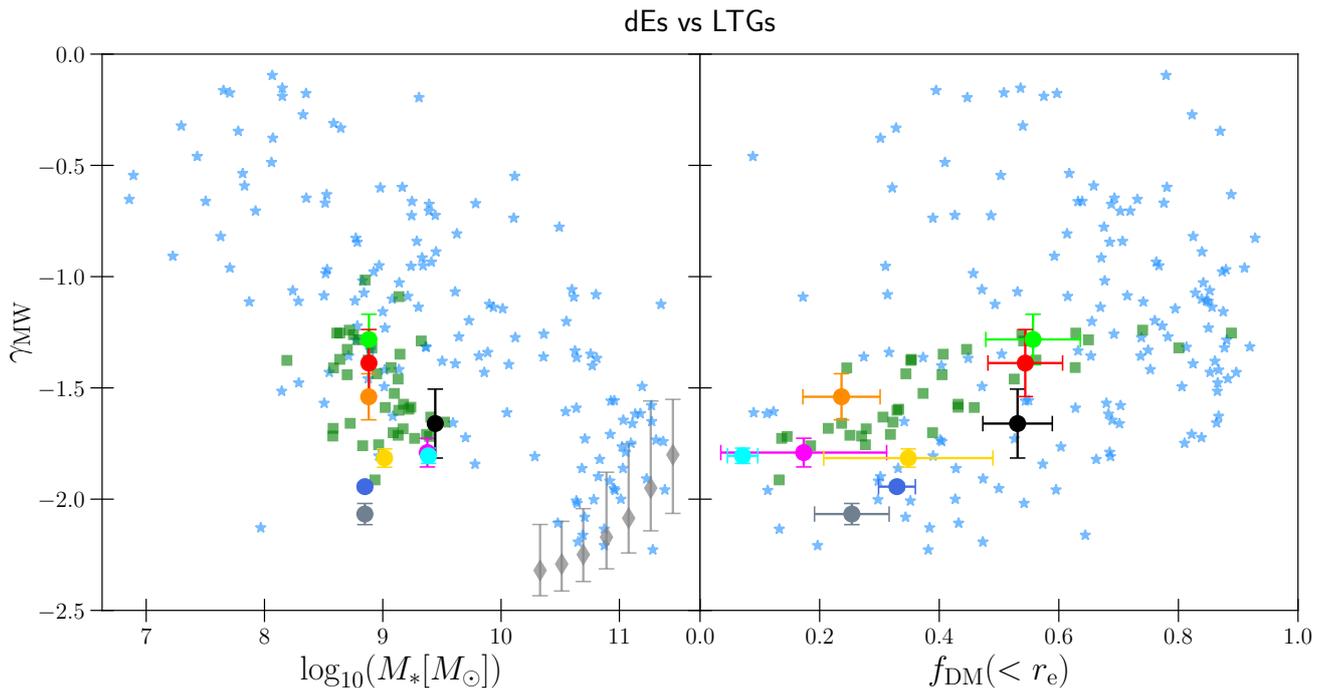


Figure 7. The mass-weighted total density slope γ_{MW} (eq. 6) of dEs and LTGs within $1r_e$ vs stellar mass M_* (left panel) and dark matter fraction f_{DM} (right panel). *Colored Dots:* Our dE sample. All the literature data shown here are from Tortora et al. (2019) who analyzed the slopes of the following 3 galaxy samples. *Blue stars:* The LTG data originate from a subsample of the SPARC survey (Lelli et al. 2016), it includes LTGs ranging from S0 to Im. *Green squares:* The dEs of the SMAKCED sample (Toloba et al. 2014; Tortora et al. 2016) which are also part of Virgo cluster. As a reference we also include the location of ‘ordinary’ ETGs in the left panel (*gray diamonds*), they stem from the ETGs of the SPIDER survey (La Barbera et al. 2010; Tortora et al. 2012). For these ETGs only the sample medians binned in different mass bins are shown, the error-bars indicate the 1σ percentiles of the more than 4000 ETGs (cf. Tortora et al. 2019).

highlight a process that may be specifically crucial in this regime alone. This transition region of the above correlations is located where the *most massive* of LTGs are found. These massive LTGs likely experienced continuous gas supply from their environment. This continued dissipative infall of baryons could be accompanied by an adiabatic *contraction* of the dark matter halo (e.g. Blumenthal et al. 1986; Gnedin et al. 2004; Li et al. 2022) which would lead to a cuspier halo and possibly also to an overall smaller γ_{tot} within $1r_e$. Since the total stellar mass traces not only a galaxy’s ability to accrete and hold on to new gas but also its past accretion history we may expect the halo contraction to be more important for massive systems, which could support the anti-correlation of the total slope and M_* in LTGs that was observed by Tortora et al. (2019); Li et al. (2019).

iii) In the *low-mass regime* (i.e. $M \sim 10^9 M_\odot$), both dEs and LTGs, have similar light profiles and slopes (e.g. dEs have nearly exponential Sersic indices (e.g. Ferrarese et al. 2006), see also stellar distributions in VW-I). The similarity in stellar profiles implies that the offset in total slope between the early- and late-type dwarfs stems

from a difference in their dark matter distributions. And this does indeed seem to be supported by the dynamically *decomposed* halo densities: the core cusp tension usually is milder in early-type dwarfs than for late-types counterparts (cf. Sec. 5.2).

In opposition to the galaxies in the intermediate mass-regime the effect of adiabatic halo *contraction* could be significantly diminished for the dwarfs. This is because as the total mass and, thus the potential well, is decreased continued gas accretion is dampened or shut-off early (e.g. by their environment) hindering the dark matter halo to become more cuspy. As a result, the dwarfs may not be able to reach the steep total slopes of ≈ -2.3 .

Conversely, to the reduced contraction adiabatic halo *expansion* may be an important factor for dwarfs that drives the anti-correlation of mass and total slope in the low-mass regime: any outflow of mass, such as a gas outflow caused by supernova feedback by newly formed stars could force an initially cuspy DM halo to become more cored as time passes (Read & Gilmore 2005; Governato et al. 2010; Pontzen & Governato 2012; de Souza

et al. 2011; Pontzen & Governato 2014; Madau et al. 2014; Read et al. 2016). The efficiency of this expansion mechanism is expected to decrease as the total mass of the systems is increased, again due to the ability of the potential well to hold mass outflows. Consequently, lower mass galaxies would become shallower than higher mass galaxies, which could produce the $\gamma_{\text{MW}}-M_*$ correlation that is seen in the low-mass regime.

In principle, this mechanism of adiabatic expansion could also explain the offset between quiescent and star-forming dwarfs that is observed in their slopes and dark matter fractions (Fig. 7). The star-forming LTGs may have had a more gradual, but prolonged, star formation history as they continuously processed their surrounding gas reservoir unaffected by the low-density environments these galaxies typically inhabit. The continued supernova outflows that result from this star formation could have gradually and gently expanded the galaxy’s size (reducing its surface density, increasing r_e and therefore also f_{DM}), while overall making the density distributions shallower (directly for baryons and indirectly for the DM distribution).

In contrast, for the quiescent dEs adiabatic expansion may have played less of a role because their star formation was being quenched at some point in their past. A manifold of potential quenching mechanisms responsible for this have been suggested (see VW–I for more discussion). Among them are *external* processes induced by the environment (e.g. ram-pressure stripping, galaxy starvation, and harassment (e.g. Gunn & Gott 1972; Larson et al. 1980; Lin & Faber 1983; Moore et al. 1998)) and *internal* processes like outflows caused by rapid supernovae or AGN feedback (e.g. Dekel & Silk 1986; Silk 2017; Sharma et al. 2023b). These mechanisms can be expected to act within different epochs and on very different timescales, for example ram-pressure stripping is expected to act on a few hundreds of Myrs to a Gyr (Quilis et al. 2000; Steinhauser et al. 2016). If the quenching mechanism that is responsible for the dEs was *fast* acting and *shortly* after galaxy formation then adiabatic expansion could have been suppressed for the dEs leaving their density distributions more cuspy and their effective radii smaller.

However, as established and discussed in VW–I (unlike the more luminous, ‘ordinary’ ETGs) many of the dEs likely had a complex star formation history, possibly involving multiple rapid star formation bursts. Depending on their initial conditions (environment and total mass) some dEs were quenched shortly after the reionization epoch while others were able to form stars up until just a few Gyrs ago. Therefore, if adiabatic expansion is responsible for the shallower densities of star-forming

dwarfs compared to the dEs, we may also expect that (within our sample) those dEs that have had prolonged SFH also are more shallow than dEs that were quenched prematurely (e.g. by a single burst). This could reveal itself in a correlation of density slopes with markers of SFH (see next section).

Given galaxies ab initio reside in halos with identical profiles and shapes, but the amount of sustained star formation evolves them in different ways we expect to find the clearest imprint of this correlation in the slopes of the dark matter (not the total density) as the halo’s slope traces the gravitational impact of mass outflows largely independent of the stellar morphology.

The dark matter fraction f_{DM} and total slopes as defined above are aperture dependent measurement because they are measured at $1r_e$. This links stellar and DM slopes indirectly and impairs an objective comparison of their halos. For example, as stated above LTGs tend to have higher r_e at the same mass, consequently, for halos of LTGs the dark matter fraction is measured at larger radii. Therefore, even if star-forming and quiescent dwarfs *currently* reside in *identical* halos, one may measure very different slopes and dark matter fractions. To eliminate these aperture-effect from the analysis, it may be preferable to analyze the dark matter halos within a physically constant aperture, which of course is only sensible if the galaxies are roughly in the same mass regime. Therefore in the following section we will again analyze the halo in the distance-independent 0.8kpc aperture we introduced in Sec. 5. This allows us to assess the dark matter densities independent of the stellar distribution they host and enable a more objective evaluation of the halos formation and subsequent evolution.

6.2. What makes dE halos cored or cuspy?

The last sections showed that dEs tend to have shallower total slopes than ‘ordinary’ ETGs, but still steeper ones than star-forming dwarfs of the same total stellar mass. To entangle whether these disparities are due to differences in the DM distributions or differences in the stellar distributions, we analyze the dynamically decomposed DM halos of our dE sample in an aperture that is independent of the stellar extent. One of the goals of this section is to explore whether the halos are initially universal but evolve over time due to their interactions with baryons and environment (Sec. 6.2.1) or, conversely, whether the initial condition of the DM halo determine the morphological evolution of the stars/baryons while the halo remains largely unchanged (Sec. 6.2.2).

6.2.1. Do the halos evolve due to internal feedback?

As considered in Sec. 6.1.3 a plausible explanation for the observed slopes are baryonic in- and out-flows (e.g. by supernova feedback). In Sec. 5.2 we discussed that the DM slopes of quiescent dwarf samples (dEs/dSphs) show considerable diversity: some are relatively cored, some are cuspier. Similarly, the periods of active star-formation are expected to vary significantly across different dEs. Depending on their initial conditions, some dEs were quenched shortly after halo formation while others likely had a complex, prolonged SFH (see VW–I or Seo & Ann 2023; Romero-Gómez et al. 2023a). If adiabatic expansion by supernova feedback is the predominant factor in evolving the DM slopes of the dEs, we expect to see a correlation of the halo properties with their stellar population properties. Assuming the dEs started out in an initially universal halo density, then the DM halos that host young, metal-rich populations (implying they were quenched recently) with a low α -abundance⁶ should be more cored and less dense in the center.

However, as evident in Fig. 8, we find only little to no correlation of the volume averaged DM densities and slopes with the stellar population parameters we derived in VW–I⁷. The dEs exhibit no correlation with metallicity and α -abundance, and therefore an extended SFH. Only the stellar age is tentatively anti-correlated with the DM slope η_{DM} . Qualitatively this age–slope anti-correlation is in agreement with the supernova feedback scenario (younger dEs were quenched recently and possibly were able to form more stars before that), however, the lack of correlation with the chemical enrichment parameters seems to indicate that adiabatic expansion by supernovae feedback only plays a minor role for the DM density distributions. Star formation seemingly does not affect the average halo density and, if any, only very moderately makes them more shallow over time. This may not be a surprise because dEs are among the most massive ‘dwarf’ galaxies which could make their halos more resilient to internal feedback, however, even the much smaller Milky Way dSphs exhibit no clear correlation with SFH either (Hayashi et al. 2020). In conclusion, if differences in SFH can not explain the diversity of DM slopes that observed in quiescent dwarfs then it

may be the case that the halos are not universal from the start (or very early on).

6.2.2. Are halo profiles and shapes universal or a product of the primordial conditions?

If the variance of DM slopes in dEs is not a product of fluctuating degrees of internal feedback, then the halos may be different ab initio (e.g. Ascasibar et al. 2004), i.e. shortly after the gravitational halo collapse but before the majority of their stars has formed. While we lack direct information about the halo structure of the dEs at high redshifts (i.e. shortly after formation) we can investigate correlations with properties that we expect to still hold information about the initial collapse conditions.

We may assume that the current position the dEs inhabit in Virgo is a reflection of their environment at halo formation. Halos that formed in the cluster outskirts may have only recently fallen into the cluster as they formed later, and/or formed in isolation within a lower background density.

Similarly, if stellar orbits in dEs were only mildly heated during their evolution (see VW–I), then the *stellar* angular momentum measured today could still hold information about the momentum of the primordial gas disk the stars formed in. And this itself is a proxy for the angular momentum⁸ of the dark matter halo (Fall & Efstathiou 1980; Fall 1983; Romanowsky & Fall 2012).

In Fig. 9 we show the correlations of the DM with the cluster environment and stellar angular momentum that we obtained in VW–I. The environment is described using a proxy parameter: the projected distance to Virgo’s central cluster galaxy M87. Galaxies that are in Virgo’s center (roughly where M87 is located) are expected to have experienced longer and stronger interactions with their environment. We either expect them to have formed very early on and enter the cluster, or they formed in-situ directly within the high density background of the cluster’s halo. We describe the stellar angular momentum using the approximate total specific angular momentum $j_* = J_*/M_*$ and the stellar angular momentum parameter λ within $0.5r_e$ (see VW–I for details on the calculation).

Conversely to the stellar population constraints, we find moderate correlations with the angular momentum and strong correlations with the environment. Galaxies

⁶ A lower α -abundance (or equivalently the $[\text{Mg}/\text{Fe}]$ we measured) is a proxy for a prolonged SFH (e.g. Romero-Gómez et al. 2023b,a).

⁷ In VW–I we derived spatially resolved populations from spectra that were binned in two annuli centered around $r = 2.5''$ and $r = 7.5''$. Shown in Fig. 8 are the results for the central aperture at $r = 2.5'' \approx 0.2\text{kpc}$. We omitted VCC 1910 because its SSP properties are not very trustworthy due to the badly constrained $H\beta$ feature (cf. VW–I).

⁸ While we measured most dEs to have near spherical halo shapes (Sec. 5) q_{DM} is likely not a good gauge for the non-radial motions of DM particles in the primordial stage: We measure q_{DM} ‘today’ (i.e. after baryonic feedback from AGNs or SF) and only within the central one effective stellar radius (the halo is much more extended and could change shape).

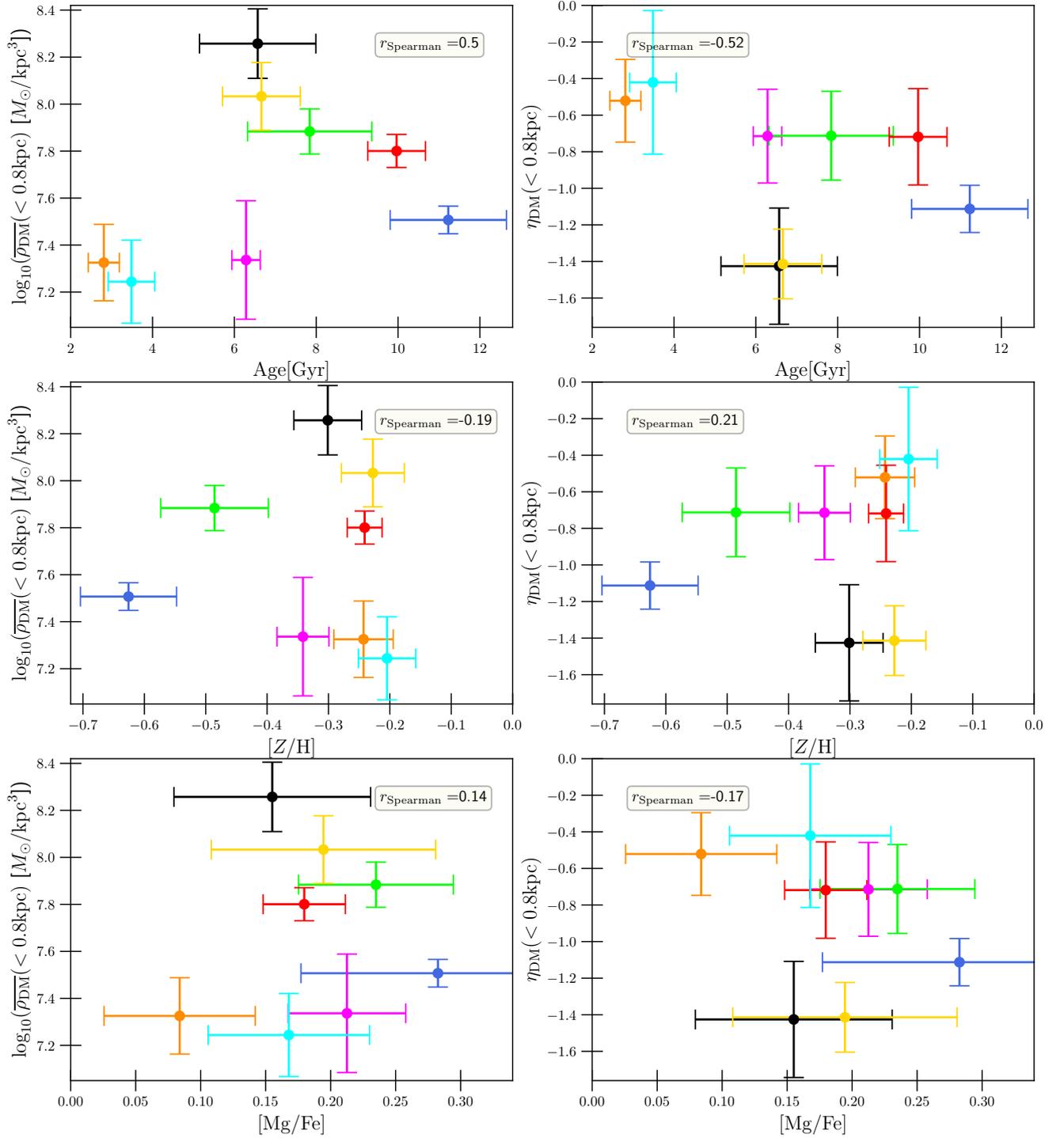


Figure 8. Correlations of the DM halos with stellar population parameters. In the small boxes of each panel, we show the corresponding Spearman correlation coefficient. *Left panels:* The average DM density $\overline{\rho_{\text{DM}}}$ (eq. 5) inside the sphere of radius 0.8kpc. *Right panels:* The average DM slope (eq. 4) inside the same sphere. *Top to Bottom:* The age of the stellar population. The metallicity [Z/H]. A low abundance ratio [Mg/Fe] is an indicator for a prolonged star formation history (cf. VW-I).

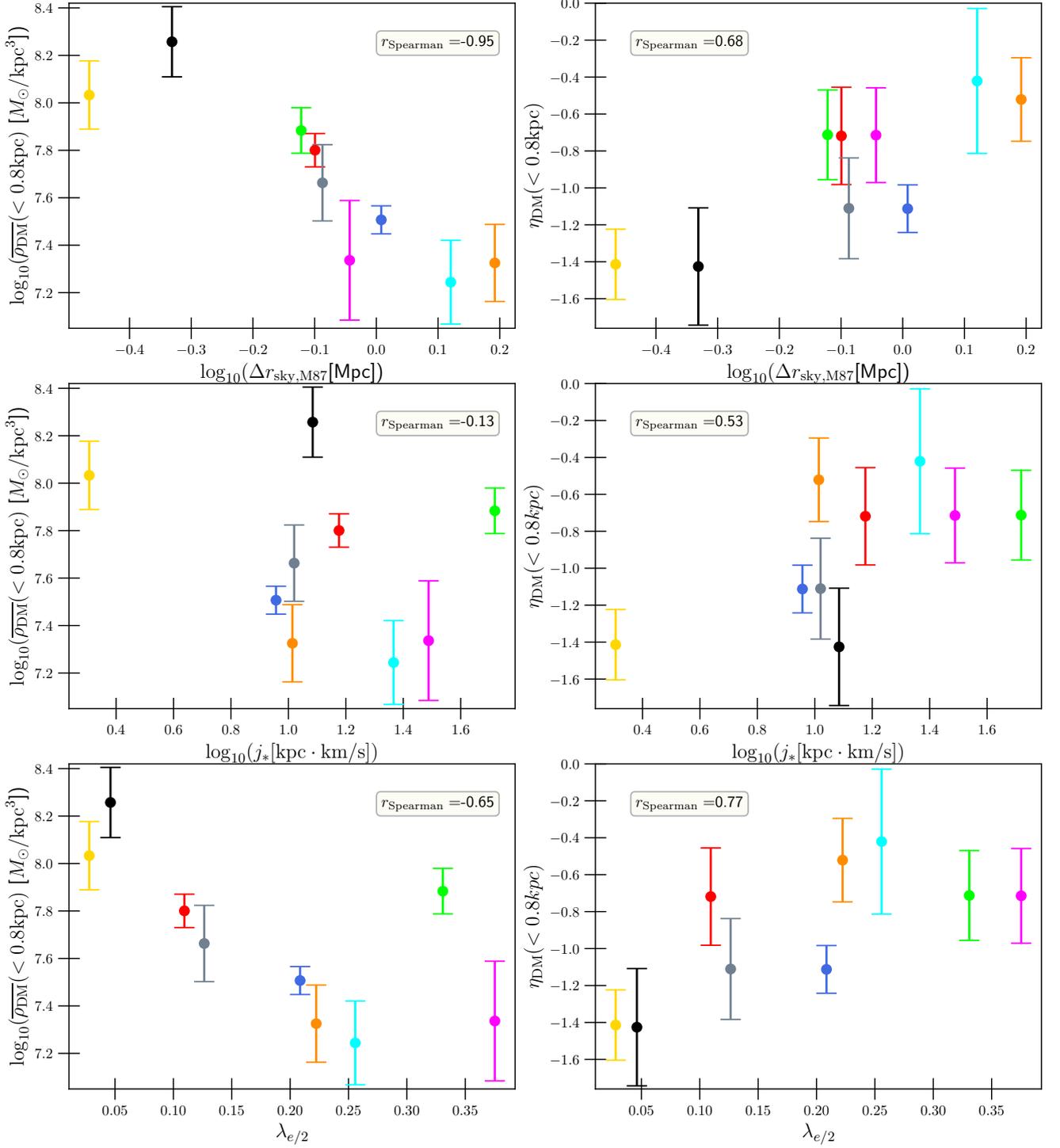


Figure 9. As Fig. 8 but for correlation of DM with cluster environment and *stellar* angular momentum. *Top to Bottom:* The projected distance to the central cluster galaxy M87 $\Delta r_{\text{sky},\text{M87}}$ (a proxy for the environment the galaxies inhabit, see VW-I). The total stellar angular momentum j_* . The stellar angular momentum parameter $\lambda_{e/2}$ (see VW-I for details) within the half the effective radius, which is not to be confused with the often used halo spin parameter (Peebles 1969).

in the cluster center have about 1dex higher DM densities than galaxies in the Virgo outskirts. Furthermore central dEs are super-NFW ($\eta \sim -1.4$) whereas galaxies in the low-density outskirts of Virgo are almost as cored ($\eta \sim -0.5$) as what is found for LTGs (typically $\eta \sim -0.3$). Similarly, the halos of galaxies that have low stellar angular momentum tend to be more dense and cuspy⁹. Only the correlation of average DM density with j_* is relatively ambiguous, but this is mostly driven by VCC 856 (green) and VCC 1261 (black)¹⁰

These angular momentum correlations may be understood as an imprint of the joint gravitational collapse of the DM and baryons. When a DM overdensity forms it acquires baryons due to its gravitational pull, these baryons condensate and infall towards the halo center, as a response the halo undergoes contraction (Blumenthal et al. 1986; Gnedin et al. 2004). However, this steepening of the halo profile during the formation phase is counter-acted by processes that make the halo more cored and less dense: i) Baryons may not be acquired smoothly, particularly in the primordial gas-rich disks the baryons fragment and form larger clumps (Immeli et al. 2004; Aumer et al. 2010; Ceverino et al. 2010). These clumps undergo dynamical friction and, in the process, heat up and expand the DM halo (El-Zant et al. 2001; Del Popolo 2009; Inoue & Saitoh 2011). ii) DM and baryons are not accreted radially but obtain angular momentum from tidal torques (Peebles 1969). Depending on the strength of these torques, a halo and its primordial gas disk will have high or low angular momentum. Halos with a higher angular momentum have been linked to a flatter density profile, lower surface density and larger extent/size in both, baryons and DM (e.g. Ascasibar et al. 2004; Williams et al. 2004; Del Popolo 2009, 2012; Kim & Lee 2013).

These mechanisms may be the cause for the correlations we found for the halo slopes with angular momentum (Fig. 9). They could also explain the cuspiest slopes

(total+DM) and smaller effective radii of the dEs when compared to LTGs (Fig. 7) since the angular momentum of dEs is suppressed in comparison to LTGs (see VW-I). Furthermore the reduced baryonic surface density that comes along with the higher halo spin could be linked to a higher observed dark matter fraction (Sharma et al. 2023a).

At first glance, this argument seems counter-intuitive when combined with the observed environment correlation. If the angular momentum grows via tidal torques then one may expect that halos in the cluster center have (on average) experienced more tidal interactions with neighboring structures and, as such, have higher angular momentum and shallower DM slopes (i.e. opposite to the observed $\eta_{\text{DM}} - \Delta r_{\text{sky}, \text{M87}}$ correlation). But perhaps the assumption that dwarfs in the cluster center should have experienced more tidal torques is an oversimplification and misses the complexity of the problem.

DM halos acquire angular momentum from tidal torques mostly *during* the linear phase of gravitational assembly (Porciani et al. 2002; López et al. 2019). During the comparatively short time of a halo's assembly epoch (i.e. at a higher redshift) the environmental circumstances may have been very different when compared to the state of Virgo at $z = 0$. If the halos assembled at very different formation epochs, their environment and, thus, their ability to acquire angular momentum may change as well. Further complications arise because halos that assembled at early epochs are expected to be denser on average due to a higher background density (see below) and because the geometry of the large-scale surroundings is anisotropic (structures grow along filaments) which affects the angular momentum acquisition of the galaxy halos (e.g. Codis et al. 2015). Perhaps the duration of halo assembly also varies with the environment and formation epoch. In a low density background DM halos may collapse on different timescales, allowing the halos to gradually acquire angular momentum.

In conclusion, while currently not obvious if and by how much, we may presume that angular momentum acquisition depends in a complex manner on a halo's assembly epoch and the geometry of its surroundings at the time. As discussed in VW-I and Sec. 6.2.1, for the dEs the derived stellar population ages are likely not a good estimator of the halo's formation epoch because many of them probably had extended periods of star formation. Therefore we rely on alternatives to gauge when the dE halos have assembled. Such an alternative is the average DM density we derived above: $\overline{\rho_{\text{DM}}}$ is a proxy for the assembly epoch of a halo because it is partly inherited from the average density of the Universe

⁹ This correlation does not necessarily mean causation, in VW-I we discussed the correlation of angular momentum with environment due to dynamical heating. Perhaps the halo profiles solely depend on the environment, but the baryons were also dynamically heated by interaction with the environment, resulting in the observed correlation of angular momentum and DM.

¹⁰ The classification of VCC 856 as a dE may be debatable as it could be a face-on spiral resulting in higher j_* (cf. VW-I). VCC 1261 is actually known to have very little to no rotation within the FoV in our data or in literature (Geha et al. 2003; van Zee et al. 2004; Chilingarian 2009; Rys et al. 2013; Toloba et al. 2014, 2015; Sybilska et al. 2017; Şen et al. 2018), but it is also the brightest dE (j_* increases scales magnitude) and has the largest effective radius which could bias the coarse approximation of the total j_* high compared to the other dEs (cf. VW-I).

at the time of assembly, as shown with collapse models and N-body simulations (Gunn & Gott 1972; Wechsler et al. 2002; Springel et al. 2005; Gao et al. 2005). Consequently, galaxies that have been assembled in an early epoch where the average background density of the Universe was higher are also expected to have denser halos than galaxies that formed during later stages of the Universe. While the proxy $\overline{\rho}_{\text{DM}}$ may be superimposed and affected by the manifold of processes (e.g. feedback from prolonged SFH, adiabatic contraction, etc.) the above discussion suggests these effects play a secondary role. They may affect the slopes to some degree, but they are unlikely to be able to reduce the DM density by $\sim 1dex$.

As evident from Fig. 9 the average density is most strongly correlated with the environment, which is very compatible with the assumption that the assembly epoch (or more generally the average background density at formation) is the primary determinant of $\overline{\rho}_{\text{DM}}$. Therefore dEs with a higher $\overline{\rho}_{\text{DM}}$ have formed in a higher density background, e.g. ‘in-situ’ within Virgo halos and/or when the Universe was young and more dense in general which gave them enough time to sink to Virgo center as seen today. Conversely, the dEs currently located in the outskirts have lower $\overline{\rho}_{\text{DM}}$ as they have formed in a lower density background (in isolation) and/or in a more recent epoch such that they only now enter Virgo proper.

In this context, the correlations in Fig. 9 suggest that dEs which have formed early (or in-situ) were less efficient in acquiring angular momentum than dEs that have formed more recently or in isolation¹¹.

Fig. 9 suggests that the average central DM density *and* the central slope are both correlated with the environment. However, $\overline{\rho}_{\text{DM}}$ and η_{DM} are not entirely *independent* properties: a steeper central slope also entails higher average central DM density and vice versa. Therefore it is not obvious that the two environment correlations with $\overline{\rho}_{\text{DM}}$ and η_{DM} are in fact two independent correlations or just an imprint of one and the same correlation combined with a correlated measurement uncertainty. We explore this possibility in Fig. 10 which shows $\overline{\rho}_{\text{DM}}$ vs η_{DM} for the 25 best dynamical models for each dE (i.e. the number of models we used to gauge the errors). As expected, the errors in $\overline{\rho}_{\text{DM}}$ and η_{DM} are correlated: if the central slope scatters to steep values, the average central density also often tends to

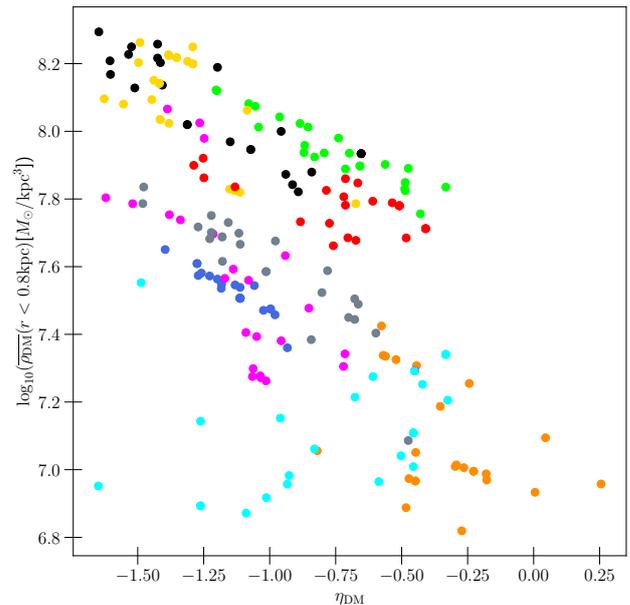


Figure 10. Correlation uncertainty of the average enclosed DM density $\overline{\rho}_{\text{DM}}$ with slope η_{DM} . Shown are the 25 best dynamical models we found for each dE.

be higher. However, overall, the scatter for a given dE is significantly smaller than the differences between the dEs in our sample. In other words, the dynamical measurements uncertainties are not enough to explain the correlations we find in Fig. 8 and Fig. 9. And we conclude that the correlations of $\overline{\rho}_{\text{DM}}$ and η_{DM} are two independent and physically meaningful findings.

6.2.3. Has the environment disturbed the halos after assembly?

In Sec. 6.2.1 we discussed the effects of *internal feedback*, but we only found tentative trends with SFH indicators. The halos may become slightly more shallow if SF is active for a long time, but this supernova feedback scenario is unable to explain the larger differences in $\overline{\rho}_{\text{DM}}$ and, consequently, the strong $\overline{\rho}_{\text{DM}}-\Delta r_{\text{sky},\text{M87}}$ correlation. Instead of a variability in primordial conditions (Sec. 6.2.2) an alternative cause for the DM-environment correlations could be tidal interactions with other cluster members and interaction with the intra-cluster medium that changed the halos *after* they assembled. This could happen by regulating star-formation (and therefore internal feedback) or directly by tidally induced mass in- and out-flows.

In principle, the $\overline{\rho}_{\text{DM}}-\Delta r_{\text{sky},\text{M87}}$ and $\eta_{\text{DM}}-\Delta r_{\text{sky},\text{M87}}$ correlations could be a result of quenching via ram-pressure stripping (RMS) due to it regulating SF. We expect prolonged SFH to gradually make halos more cored and reduce $\overline{\rho}_{\text{DM}}$ (Sec. 6.2.1), but RMS stops this

¹¹ In terms of stellar population age VCC 200 (dark blue in Fig. 8) has formed shortly after reionization (12 Gyrs) yet has moderate $\overline{\rho}_{\text{DM}}$, therefore we may suspect that the local background at formation (in-situ or isolation) and not epoch plays the leading role.

process. If the dEs in Virgo’s outskirts were only recently being quenched by RMS as they entered Virgo they may have had more time between assembly and quenching to hollow out their halo, producing the correlations with $\Delta r_{\text{sky}, \text{M87}}$. In this scenario it is not the process itself (RMS) that drives this but the ceasing of internal feedback¹². However, we found little evidence that internal feedback is particularly important (Sec. 6.2.1). Furthermore we may expect a significant fraction of dEs to be quiescent even before entering Virgo for the first time due to pre-processing in the groups they arrive in (Bidaran et al. 2022; Romero-Gómez et al. 2024). Therefore, if SFHs impacts the dE halos, we would expect milder halo-environment correlations as those group dEs would have retained their DM distribution.

Unlike ram-pressure stripping, we may suspect other environment processes to affect the slopes of the dark matter more directly. Examples include tidal harassment by larger galaxies, and mergers with galaxies of smaller or similar size. Galaxy harassment removes both stars and dark matter from the potential well and dynamically heats their orbits. However, the dark matter correlations (Fig. 9) stand in exact opposition to this scenario, since tidal removal would be expected to reduce the dark matter density and flatten the halo.

Galaxy mergers violently change the mass structure and orbits of dark and baryonic matter. Wet mergers are unlikely in the cluster’s center, but could play a role before the progenitors of the dEs enter Virgo by inducing increased star-formation and rapid bursts. As is the case with tidal harassment though, this effect would be converse to our finding of more dense/cuspy halos in Virgo’s center where the dry merger-rate is expected to be higher¹³.

Altogether the direct impact of the environment on the dEs halos *after* assembly should be negligible, if anything it affects the halos indirectly by regulating their star formation and as a consequence internal feedback.

6.3. The formation of dEs - Are dEs transformed remnants of spirals?

¹² While RMS may displace dark matter (Smith et al. 2012) in the initial quenching phase as the outflowing gas is being stripped, its long-term impact on the slopes and halo density is probably low. Since all dEs are quenched by now, RMS would have affected all dEs equally and an environment correlation wouldn’t remain.

¹³ There are some simulations (for less massive dwarfs) that suggest that dry *major* mergers can sometimes make halos *cuspy* by importing additional dark matter to the center (Laporte & Penarrubia 2015; Orkney et al. 2021) but this depends on initial conditions and would likely be a stochastic effect unable to produce the strong environment correlations.

The synthesis of Sec. 6.2 is that the diversity in the halo distribution of dEs are largely a result of the primordial conditions during halo assembly, but modified (to a lesser degree) by the secular evolution caused by the internal SF-feedback the galaxy was able to maintain before it was quenched. By comparing the halos of different morphological types with one another, we may be able to infer how dEs have arrived in their current quiescent and homogeneous stellar state (cf. VW-I).

The prevalent formation scenario is that dEs are the remnants of late-type progenitors that were, at some point in their evolution, quenched by some internal or external process (see discussion Sec. 1 and VW-I). However, the results discussed in Sec. 6.2 suggest that the DM (especially $\overline{\rho_{\text{DM}}}$) is unaffected by the quenching mechanism itself and only mildly affected by feedback from its SFH. Consequently, if dEs have formed from late-type progenitors, we would expect their DM structure to be relatively robust and comparable to that of spiral dwarfs that were able to avoid quenching. In other words, differences in SFH may have made the LTG dwarfs slightly more cored on average (see Sec. 5.2) but we would expect $\overline{\rho_{\text{DM}}}$ to be similar as it is mostly determined by the background density during halo assembly.

Fig. 11 shows the average DM density of our dEs compared to galaxies of various morphological types: ‘ordinary’ ETGs in the Coma cluster (Thomas et al. 2009), dSphs of the Milky Way (Burkert 2015), and the presumed progenitors of dEs: Sc-Im galaxies (Kormendy & Freeman 2016). Since we now compare galaxies spanning a large range of magnitudes and effective radii, we have opted to calculate $\overline{\rho_{\text{DM}}}$ within $1r_e$ instead of the size-independent aperture we used in the last sections. To compare the different samples consistently, we decided to use the total, extinction-corrected luminosities in the B band throughout.

Similar scaling relations of the average (central) DM density with luminosity have been investigated before and both, ETGs and LTGs, were found to have DM densities that are anti-correlated with luminosity, though the scaling relation for the ETG population is offset by $\sim +1\text{dex}$ suggesting an earlier halo assembly consistent with their older stellar populations (cf. Gerhard et al. 2001; Thomas et al. 2009). In contrast, the quiescent dSphs of the Milky Way fall on the extension of the scaling relations of larger LTGs (Kormendy & Freeman 2016), i.e. they are compliant with the formation scenario that they are the remnants of dIm galaxies. These previous results are reproduced in Fig. 11 as indicated by the dashed lines that are fitted scaling relation to the ETG and Sc-Im galaxies. While the luminosity range of our small dE sample is too narrow to establish a similar

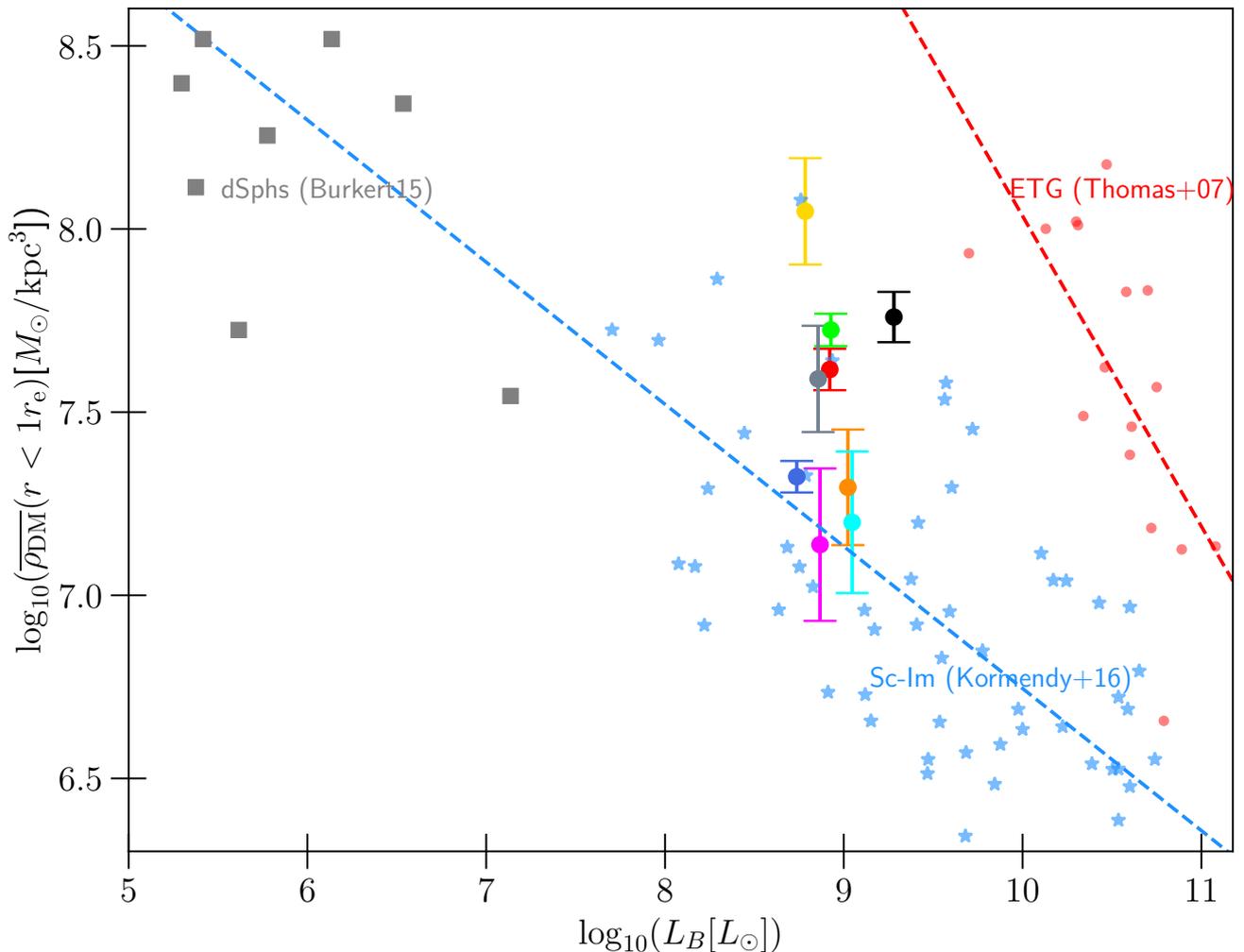


Figure 11. The B-band Luminosity vs average DM density within an aperture of 1 effective radius for different galaxy morphologies. *Colored with errorbars:* Our dE sample. The effective radii we used to calculate $\overline{\rho_{\text{DM}}}$ are in the z-Band and tabulated in VW-I. *Red points:* Large ETGs in the Coma cluster as measured by Thomas et al. (2009). Following Thomas et al. (2009) we obtained their effective radii from Jorgensen et al. (1995); Mehlert et al. (2000) and assumed a Coma distance of 100 Mpc. *Yellow points:* Sc-Im galaxies from Kormendy & Freeman (2016), this sample includes galaxies in different groups within the local volume of 84 Mpc. *Gray squares:* The dSphs of the Milky Way from Burkert (2015). *Dashed Lines:* For the LTG and ETG samples we also show fitted scaling relations in the same color. For the samples of Burkert (2015) and Kormendy & Freeman (2016) we plot the *core densities* as they are stated in these studies instead of recalculating the averaged density within the $1r_e$ aperture. This is a good approximation of $\overline{\rho_{\text{DM}}}$ because these studies all use halo models with a central core (with minor differences in the model definitions) but find core radii that are considerably larger than one stellar effective radius. Therefore, if we were to correct for the small radial decrease within $1r_e$ it would shift the density only slightly to lower values which would not change our conclusions (see text). Between different studies we expect the systematic differences/uncertainties in the measurement of r_e , and in the probed candidate halo models (Sec. 3) to have more of an impact. The dEs do not follow the scaling relations suggested by large ETGs. While they are closer to the DM densities of Sc-Im galaxies, which (presumably) are the progenitors of dEs, they are offset to higher densities. This could indicate that the progenitors of dEs have assembled earlier and/or in a higher density background.

scaling relation for dEs we can compare our galaxies to the existing scaling relations.

The $\overline{\rho_{\text{DM}}}$ scatter within our dE sample stays within 1dex which is comparable to the scatter that was found within the individual literature samples. As argued in Sec. 6.2 we deem this scatter to be the genuine variety in halo distributions with assembly environment, age and feedback. The dEs follow neither the LTG nor the ETG scaling relations well. The extension of ‘ordinary’ ETGs would predict much higher densities for the dEs. The comparison with the presumed progenitors of the dEs, the Sc-Im galaxies, is less obvious. While they all fall inside the distribution of LTGs of similar luminosity, the dE sample average is offset higher by $\sim +0.5\text{dex}$. This 0.5dex offset is based on a comparison of LTGs and ETGs at the *same* luminosity L_B . But this does not account for the fact that LTGs will also be offset towards lower stellar mass-to-light ratios due to their younger populations. Consequently, if we were to compare the LTGs at the *same* stellar mass, this offset between dEs and Sc-Im would only increase further. Therefore a comparison at the same stellar mass would increase the following conclusions, however, considering the luminosity is a more straight-forward measure and not depend on correct mass-to-light ratios, we decided to compare versus the luminosity in Fig. 11.

Generally speaking, we expect the position of a galaxy in the $\overline{\rho_{\text{DM}}} - L_B$ -space to be determined by the assembly conditions (i.e. formation epoch, environment, angular momentum, ...) and the subsequent evolution (e.g. SF-feedback, tidal interaction, ...). Since baryons and dark matter are usually lost/expelled by the latter (e.g. Dekel & Silk 1986) such evolutionary processes shift galaxies diagonally to lower $\overline{\rho_{\text{DM}}}$ and L_B , i.e. it could move dEs closer to the LTG scaling relations. However, our results suggest that, at least in the mass regime of our dEs, $\overline{\rho_{\text{DM}}}$ is only marginally affected by the latter (Sec. 6.2) despite the fact that some of the dEs seemed to have formed stars actively over a very prolonged period. For the dEs to fall onto LTG relations much more feedback would be needed and it is not obvious why Sc-Im galaxies in the same luminosity/mass regime should not be affected similarly to the dEs.

The stripping scenario of dEs can, in principle, ‘fake’ higher average DM densities for dEs within $1r_e$ as we may expect it to change the stellar distribution (and indirectly the DM distribution) and thus affect the aperture we measure ρ_{DM} in. To exclude this possibility, we can compare the DM density in the very center of the halo $\rho_0 = \rho_{\text{DM}}(r=0)$. This is because we expect the central density to be less susceptible to external influences and systematic aperture differences between the

LTG and dEs galaxy types. As Fig. 12 shows the offset of the dEs becomes even larger when comparing the central DM densities which is no surprise because comparatively the Sc-Im have relatively cored halo profiles. This confirms external influences like stripping are not able to explain the *higher* DM density of the dEs. Therefore it is not obvious how an Sc-Im galaxy with a lower ρ_0 could be transformed to a dE with a higher ρ_0 .

Considering all of the above reasons, we suspect the offset between Sc-Im and dE galaxies is a genuine difference in their DM density. In that case, the offset suggests two physical (and likely correlated) reasons: i) dEs have assembled in a higher density background, because they were formed at a higher redshift and/or in-situ in a cluster’s dense environment. ii) Internal feedback (Supernovae, AGNs, etc.) was different (stronger) in the LTGs we see today which has shifted them away from the dEs¹⁴. If LTGs were progenitors of dEs this would imply that the strength of internal feedback must have increased the later a galaxy has assembled, which poses the question of why LTGs seen today (which reside in lower density halos) were not self-quenched yet¹⁵.

6.4. Distinct formation channels of today’s quiescent and late-type dwarfs.

In either case, i) or ii), the Sc-Im galaxies observed today are not directly comparable to the supposed progenitors of dEs because $\overline{\rho_{\text{DM}}}$ can not be increased after assembly. Many theories of galaxy evolution attempt to explain the links/similarities between late-type dwarfs and dEs using processes (e.g. RMS stripping or SF feedback) that transform $z=0$ late-types into $z=0$ dEs. But as already stated by Skillman & Bender (1995), perhaps this is “asking the wrong question” as such theories assumes that the progenitors of dEs are comparable to today’s Sc-Im galaxies. Often this makes these theories struggle to explain the co-existence of dEs and Sc-Im within the local Universe. Combined with the findings of VW-I, which suggest that dEs have a suppressed stellar angular momentum compared to $z=0$ LTGs in the same mass regime, we hypothesize two distinct formation channels for dEs and late-type dwarfs observed in the $z=0$ Universe.

¹⁴ The direct impact of external/tidal effects must also be negligible because it would have affected the dEs more than the more isolated LTGs (Dressler 1980). This is congruent with our findings in Sec. 6.2.3.

¹⁵ If dwarfs in this luminosity regime are generally not self-quenched, the dEs we observe must have been quenched solely by the environment (cf. VW-I) but this is challenged by the existence of isolated dEs in the field (e.g. Janz et al. 2017; Paudel et al. 2023).

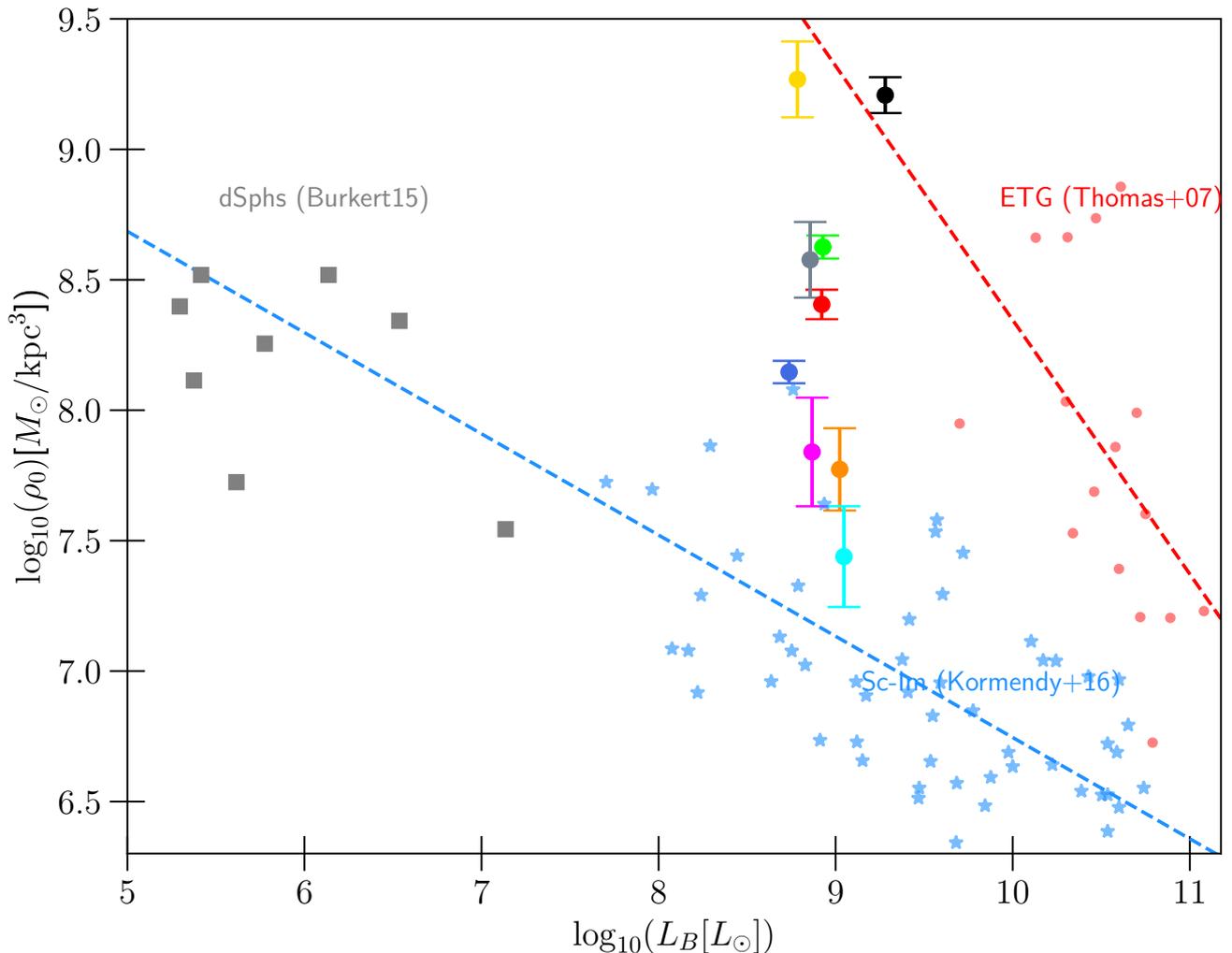


Figure 12. As Fig. 11 but for the *central* DM density ρ_0 (i.e. the maximum of the DM density). For the dEs ρ_0 is measured as the average density within the central 100pc, for the massive ETGs (Thomas et al. 2009) within 1kpc. The dSphs and Sc-Im galaxies remain unchanged relative to Fig. 11 because in both figures we show the parameter $\rho_0 = \rho(r=0)$ of the cored halo profiles that these studies employ to model the halos.

The progenitors of present-day dEs assembled at a higher redshift within a much denser environment in general. In these extreme conditions, angular momentum acquisition may have been suppressed and/or star formation was very different from today (e.g. higher efficiency and very bursty because of the higher avg. density). The former could have made the total density slopes (Fig. 7) steeper, as baryonic contraction is more efficient. The latter could have led to some of the dEs being self-quenched after a single rapid SF burst (e.g. VCC 200) while others were just massive and isolated enough to be able to re-accrete some of the expelled gas (rejuvenating SF) until they were eventually quenched when they first entered the cluster (see VW-I). Additionally, under the extreme early assembly conditions

we may expect AGN activity to become more important than in dwarfs that formed recently. AGNs could quench galaxies, affect star-formation and affect the dark matter dynamically (e.g. Koudmani et al. 2022; Arjona-Galvez et al. 2024).

Conversely, the $z=0$ late-type dwarfs have all assembled more recently in an epoch where halo assembly happens under different circumstances and SF is prolonged but less bursty (avoiding self-quenching). Their lower average DM density makes them more susceptible to external forces like harassment, which could explain the lack of dwarf late-types in the cluster centers. These distinct formation channels of today’s quiescent and star-forming dwarfs could also resolve why the latter have higher DM fractions (Sec. 7) than the dEs. The

first JWST observations of high-redshift dwarfs ($z \gtrsim 6$, i.e. the progenitors of dEs in this scenario) and corresponding simulations (de Graaff et al. 2023, 2024) suggest that galaxies in this mass-regime have initially high dark matter fractions $f_{\text{DM}} \sim 0.8$ but evolve to $f_{\text{DM}} \sim 0.4$ at cosmic noon as they form more and more stars. Perhaps the star-forming galaxies at $z = 0$ are also still within a dark matter dominated phase, as they are still in the process of converting their gas into stars.

Still, there are caveats to the conclusions drawn from Fig. 11: i) In our work, we have probed halo models that are very flexible (Sec. 3) allowing for a variety of profiles and shapes. Many studies use more restricting assumption about the halo profiles (e.g. spherical NFW- or cored profiles) which could bias ρ_{DM} . We plan to explore the effects of model choice in LT. ii) $\overline{\rho_{\text{DM}}}(< r_e)$ is not an aperture independent quantity. At the same stellar mass the effective radii of LTGs may be systematically larger than that of a dEs with the same stellar mass. Then, if we assume an LTG and a dE sit in an identical halo one measures a smaller $\overline{\rho_{\text{DM}}}$ for the former. We estimate we would have to increase the effective radii (i.e. the aperture) of our dEs by a factor of ~ 2.1 without changing the halo densities in order for them to fall nicely on the LTG scaling relation. iii) We can not easily distinguish between a difference in environment and assembly epoch yet compare galaxies of different groups and clusters, e.g. one cluster/group could have formed later than another.

Virgo is believed to be a dynamically young cluster. It has a loose, cross-shaped structure with several intact sub-clumps (Binggeli et al. 1987) and an irregular distribution of its intra-cluster medium (Böhringer et al. 1994). Furthermore Virgo’s members exhibit a non-gaussian, unrelaxed velocity distribution (Conselice et al. 2001) and a surprisingly high fraction of LTGs for a cluster (Sandage & Binggeli 1984). The strong $\overline{\rho_{\text{DM}}}$ - $\Delta r_{\text{sky}, \text{M87}}$ correlation is also telling about the dynamical state of the Virgo cluster. As a cluster relaxes and aggregates nearby galaxy groups, these infalling groups are expected to lose most of their smaller members within a few Gyrs to the larger cluster body (cf. Choque-Challapa et al. 2019). This process would gradually break down any $\overline{\rho_{\text{DM}}}$ - $\Delta r_{\text{sky}, \text{M87}}$ correlation as the smaller members mix with the cluster. The fact that the $\overline{\rho_{\text{DM}}}$ - $\Delta r_{\text{sky}, \text{M87}}$ correlation is still this strong supports this view of Virgo as a dynamically *young* cluster.

7. SUMMARY AND CONCLUSIONS

This work presents the first dynamical analysis of a dE sample ($\log_{10}(M_*/M_\odot) \in [8.5, 9.5]$) which employs orbit-superposition modeling and resolved, higher-order

LOSVD information to constrain the detailed structure of their DM halos. One of the advantages of this approach is that it does not impose a priori restrictions on the stellar anisotropy structure, allowing us to lift the mass-anisotropy degeneracy and measure the halo profiles and shapes unbiased. The models incorporate stellar mass-to-light ratio gradients (i.e. allowing for spatial variances of stellar populations) and explore a large variety of possible halo profiles. For the first time, we also constrain the flattening of dark matter in dEs. We investigate whether the observational constraints are in tension with the standard Λ CDM paradigm (cusp-core problem and sphericity) and, if not, what they imply regarding halo assembly, star formation, environment feedback, and the progenitors of dEs. Our main conclusions are:

dE halos are only mildly cored: We do not find evidence for a strong tension of the DM slopes of dEs with Λ CDM predictions. The average DM slope of our sample is ~ -0.9 . While the dEs exhibit some diversity with slopes $\in [-1.4, -0.5]$ this is still within the range of uncertainties in baryonic physics. These observations are formally inconsistent with DMO Λ CDM simulations, but we do not necessarily need to invoke exotic physics as simulations that include baryonic feedback appropriately may be able to explain the diversity and modest preference to cored profiles.

Halos are very round: The tension with Λ CDM may be more troubling in terms of the sphericity of halos. Similar to spherical shape constraints found for the Milky Way’s halo, the dark matter of dEs is distributed nearly spherical ($q_{\text{DM}} \sim 0.9 - 1.0$) independently of the shape of the stellar distribution. While the inclusion of baryonic feedback in Λ CDM simulations is known to make halos more spherical, easing some of the tension, the constraints we measured still exceed expectation values from simulations. Still, unlike slope measurements, few constraints on the flattening exist, and the field is still in its infancy. More galaxies will need to be investigated.

Is baryonic feedback enough to explain the remaining tension? The jury is still out. It may be surprising that the dE halos seem more in tension in terms of their sphericity than they are in terms of slopes, considering the same baryonic mechanisms are used to explain both: halos becoming more spherical and cored. Simulations (Λ CDM and more exotic physics) often attempt and are successful in explaining individual aspects in isolation (e.g. why are halos cored). The challenge remains if simulations can explain multiple aspects at once. Only then we may hope to narrow down the under-

lying mechanisms if we consider all/more observational constraints simultaneously.

dE centers are dominated by luminous matter: We find upper limits for their central supermassive black hole mass ($M_{\bullet} \lesssim 10^6 M_{\odot}$) and dark matter fractions within one stellar effective radii that are moderate. The bulk of dEs have $f_{\text{DM}} \in [0.2, 0.5]$ which is slightly more than what is typical for massive ETGs but considerably less than what is found for most LTGs.

Total density slopes are shallower than those of massive ETGs but still steeper than those of LTG dwarfs: The slopes of the total density (baryons+dark matter) in dEs are relatively shallow with $\gamma \sim -1.5$. This places them on the extended anti-correlation found for stellar mass and total slope of ‘ordinary’ ETGs with $M_{\odot} \lesssim 5 \cdot 10^{10} M_{\odot}$. Therefore dEs are shallower than ‘ordinary’ ETGs who have approximately isothermal slopes, but still steeper than LTGs of the same total stellar mass.

Halo densities and profiles are strongly correlated with environment, and moderately with stellar angular momentum: In the cluster center the DM halos are denser and cuspiest than an NFW profile. Conversely, in Virgo’s outskirts, the average halo density is reduced by ~ 1 dex and the profiles are moderately cored (~ -0.5). Similarly, dEs with lower stellar angular momentum (and presumably low DM angular momentum) have cuspiest halos.

Halos of dEs evolve only mildly, and are not universal: We find these correlations are difficult to be explained by tidal interactions and quenching mechanisms like supernova feedback. While tidal interactions and ram-pressure stripping are likely important quenching mechanisms (cf. VW–I), they have little effect on the halos. Tidal interactions with the environment are incongruent with the observed correlations. Internal feedback by supernovae winds is a second order effect that may change the slopes of those dEs that have an extended SFH, but even then only moderately. In conclusion, the dE halos have evolved only mildly after their gravitational assembly.

This lack of DM evolution due to internal or external feedback, the diversity in measured DM slopes $[-1.4, -0.5]$, and the correlations with environment/angular momentum suggest that halo profiles do not have a universal shape after assembly. Depending on the conditions during their halo formation epoch (environment, angular momentum, baryon clumpiness, AGN activity, etc.) some halos will turn out to be more cuspy than NFW if baryonic contraction prevails, or more cored if counter-acting process (e.g. dynamical friction, tidal torques) dominate.

A different formation? dEs may have assembled at higher redshift in more extreme conditions than local star-forming dwarfs: Our results suggest that the dE halos in Virgo have formed at a higher redshift than LTG dwarfs of similar mass. During this early assembly epoch star-formation, AGN activity, and environment/formation conditions may have been much more rapid and extreme than for the LTGs that formed more recently. This may have led to additional mechanisms that were able to quench some of the dEs shortly after formation. Considering these extreme early stages, local spirals which seemed to have formed in a much more quiescent epoch of the Universe are not necessarily representative of the star-forming progenitors of dEs.

Our stellar population analysis in VW–I has left open two plausible scenarios: i) dEs have been continuously produced throughout Virgo’s history, but the IMF varies with the formation age. ii) dEs have been formed very early in Virgo’s history but depending on their mass and environment some of them were quenched early while others have been able to hold onto (or re-accrete) their gas, thus, experiencing a more complex SFH. The results presented here make scenario ii) more likely. This means that if the IMF varies, then it probably varies less than implied by taking the stellar masses of single-stellar-population models at face value (see VW–I). However, whether an early formation epoch is enough to make the dEs’ dynamical stellar masses compatible with a Kroupa or even sub-Kroupa IMF remains to be checked with more sophisticated stellar population models.

Future observational constraints for local LTGs could aid in further narrowing down the formation scenarios. For example, we expect that assembly at higher redshift resulted in rounder halos (Chua et al. 2019) than at $z = 0$. While our dEs are indeed surprisingly round, not enough constraints on the flattening of local LTGs dwarfs exist to compare the dEs to.

Our comprehension of small-scale cosmological problems like the cusp-core problem is inextricably linked to our understanding of galaxy formation and evolution. Galaxies are unique objects with their own history of assembly and evolution, resulting in different DM distributions. Our results suggest that there are considerable differences between dEs and comparable late-type dwarfs in the local Universe, but many more galaxy classes exist which may even be more divergent. Ultra-diffuse galaxies, for example, are also quiescent and close in magnitude to dEs, but they are very extended, have a lower surface brightness, and some of them appear to have essentially no dark matter while others have much higher dark matter fractions than the dEs (e.g.

van Dokkum et al. 2016; Danieli et al. 2019; Bar et al. 2022; Zöllner et al. 2023). Precise and accurate dynamical decompositions (App. A) of dark matter and baryons for a manifold of different galaxy types is needed, and a consistent comparison between them (e.g. in physical units instead of aperture-dependent) will be essential to investigate small-scale cosmology in the future.

ACKNOWLEDGEMENTS

This work is based on observations obtained with the Harlan J. Smith Telescope at the McDonald Observatory, Texas. Computing has been carried out on the COBRA and RAVEN HPC systems at the Max Planck Computing and Data Facility (MPCDF), Germany. We also made frequent use of the NASA/IPAC Extragalactic Database (NED), operated by the Jet Propulsion Laboratory and the California Institute of Technology, NASA’s Astrophysics Data System bibliographic services, and the HyperLeda database (Paturel et al. 2003).

Facilities: Smith, HST

Software: astropy (Astropy Collaboration et al. 2022)

APPENDIX

A. HOW WELL CAN WE RECOVER DARK COMPONENTS? - A STRESS TEST

The primary goal of this paper is to investigate the shapes, slopes, and DM fractions of the halos of dEs by use of dynamical orbit modeling. This requires a reliable decomposition of the dark components (black hole and DM halo) from the stars. As preparation for this study, we stress-tested our observational and modeling setup on an N-body simulation. The goal of this test was three-fold: i) Explore the general ability of the orbit models and our observational setup to constrain the mass and kinematic distribution of the dE sample. ii) Investigate how reliable one can decompose the individual matter components. iii) Test whether we can constrain the *flattening* of dark matter halos and gauge how much the assumption of axisymmetry could bias our results.

In Sec. A.1 we explain how we generated the mock observations for our modeling, then we continue with a general evaluation of the mass recovery and decomposition quality (Sec. A.2). We also gauge how much a possible triaxiality of the galaxies could distort the results derived with axisymmetric modeling, specifically the flattening of the dark matter halo (Sec. A.3).

We expect two circumstances to negatively affect the dark matter recovery: i) If the *local* dark matter contribution to the total mass is negligible, the recovered halo distribution is poorly constrained in these regions because it does change the gravitational potential significantly. ii) If the dark matter follows the exact same

distribution as the luminous mass distribution (a mass-follows-light distribution) the two components are degenerate since it’s impossible to dynamically differentiate between dark stellar remnants and genuine dark matter. From an orbit’s perspective, one component can be absorbed by the other (e.g. by up-scaling the stellar mass-to-light ratio) without changing the combined potential.

Point i) will be explored in detail in LT but we may account for it for now by trusting the dark matter recovery in regions where the models suggest a significant dark matter fraction more than where it doesn’t. While a reliable decomposition is completely impossible in the extreme case of point ii) we stress tested our modeling procedure by applying it on an N-body simulation that represents a tough case where at least some of the dark matter distribution follows the luminous matter rather closely within the center of the simulated galaxy. At larger radii the two mass components start to differ in their spatial distributions and amount, which, in principle, should allow the models to break the degeneracy.

A.1. The simulation setup

As a test for our modeling routines, we use the results from an idealized N-body simulation of merging low mass galaxies similar to Partmann et al. (2023), taking into account the gravitational dynamics of dark matter, stars and black holes.

The N-body simulations are run with the Ketju code (Rantala et al. 2017, 2018; Mannerkoski et al. 2021), a combination of the Gadget-3 tree gravity solver with accurate regularized integration (Springel et al. 2005). This method allows resolving the unsoftened forces between the star/DM components and black holes for an accurate treatment of dynamical friction and scattering of DM and stars by black hole binaries or multiples. The coalescence of black holes by gravitational wave emission is followed with Post-Newtonian corrections up to order 3.5.

The simulations follow the merger of five small galaxies (each with DM mass $4 \cdot 10^8 M_\odot$ and stellar mass $4 \cdot 10^7 M_\odot$) with a five times more massive central halo (DM mass $2 \cdot 10^9 M_\odot$ and stellar mass $2 \cdot 10^8 M_\odot$). Each of the galaxies consists of a DM halo (Hernquist profile with $r_{1/2} = 7.1 \text{ kpc}$), a stellar component (Hernquist profile with $r_{1/2} = 1 \text{ kpc}$) and a black hole. By construction, the six galaxies merge within the first few Gyr and result in a galaxy with a total mass of $M_{\text{tot}} = 4.4 \cdot 10^9 M_\odot$, where $M_{\text{DM}} = 4 \cdot 10^9 M_\odot$ are contributed by dark matter and $M_{\text{star}} = 4 \cdot 10^8 M_\odot$ by the stellar component. In the state we ‘observe’ the system, the central galaxy has experienced several mergers in its past (7-10 Gyrs ago) and is now de facto in an equilibrium state again.

Each of the merger progenitors carries a black hole. While Partmann et al. (2023) explores various black hole masses and merger orbits, for this study we choose an extreme case with a central black hole mass of $10^7 M_\odot$ and $2 \cdot 10^6 M_\odot$ black holes in the five infalling smaller galaxies. For reasons explained later, highly overmassive black holes can produce density cores with similar dark matter and stellar densities in the galactic center. We choose this scenario because it is expected to be the most challenging for our modeling pipeline. The extent and mass are typical for the dEs we observed in our Virgo sample. However, beyond that we do not expect the N-body simulation to be a very physical representation of the circumstances in real dEs, i.e. we do not expect a mass-follows-light distribution in dEs nor that they have experienced several dry mergers with very massive black holes.

During the merger process, most black holes sink to the center of the central halo where they have complex evolution paths that result in the formation of black hole binaries or triples, the dynamical ejection of black holes or black hole mergers. In the simulation considered here, three among the five black holes that were brought into the central galaxy with the infalling smaller galaxies were ejected by dynamical interactions. As a result, black holes with a total mass of $1.4 \cdot 10^7 M_\odot$ re-

main in the center of the central galaxy. The system of black holes in the center has a sphere of influence of $r_{\text{SOI}} \sim 0.3 \text{ kpc}$, where r_{SOI} is defined as the radius at which $M_*(r < r_{\text{SOI}}) = M_\bullet$. This M_\bullet is barely resolved as r_{SOI} is close to the spatial resolution ($\sim 0.25 \text{ kpc}$) of the observational setup of the real dE sample that we adopted for this simulated test. As such we do not necessarily expect to find strong constraints towards lower black hole mass and possibly only an upper limit for M_\bullet .

In massive early type galaxies “black hole binary scouring”, i.e. the ejection of stars from the galactic center through slingshots with a black hole binary is an important process that can convert an initially cuspy stellar density profile into a profile with a flat density core (e.g. Kormendy & Bender 2009; Kormendy et al. 2009; Kormendy & Bender 2013; Thomas et al. 2014; Rantala et al. 2018; Mehrgan et al. 2019). As discussed in Partmann et al. (2023), even in low mass galaxies the combined effect of black hole sinking through dynamical friction, black hole binary scouring and black holes ejections can lead to the formation of large stellar and dark matter density cores if the black hole masses are large enough. For the large black hole mass used here, this effect leads to the formation of a large core with a size $r_c \sim 0.4 \text{ kpc}$ (or about $5''$ in our mock setup) and an *assimilation* of the *central* dark and baryonic matter distributions where both matter components follow approximately the same density distribution with a ratio of ~ 0.5 of dark and baryonic matter. This means the two components are approximately indistinguishable within the core. Only at larger radii where the scouring had less impact on the mass distribution the two matter components begin to diverge with the dark matter starting to dominate as the stellar density declines more steeply than the dark matter. The mass-to-light ratio of the single stellar population is a spatially constant $\Upsilon_* = 1.0$, which means the total *dynamical* mass-to-light ratio is $\Upsilon_{\text{dyn}} \sim 2$ within the scouring core. Υ_{dyn} is reaching ~ 5 (or a dark matter fraction of $\sim 70\%$) near the edge of the mock field of view (FoV) at $r_{\text{FoV}} = 20''$.

Since the galaxy in the N-body simulation is a result of several mergers, the density distribution is not expected to follow a particular symmetry, hence the axisymmetric orbit models we employ are unlikely to be a perfect fit. If we approximate the particle distributions using a *triaxial* ellipsoid we find that, both the baryonic and dark matter component, follow a very similar shape profile with both components being triaxial and oblate outside the core. However, within the core the shape of both mass components becomes near spherical. Within the core the particle distribution is also more complex and neither an axisymmetric nor a triaxial approxima-

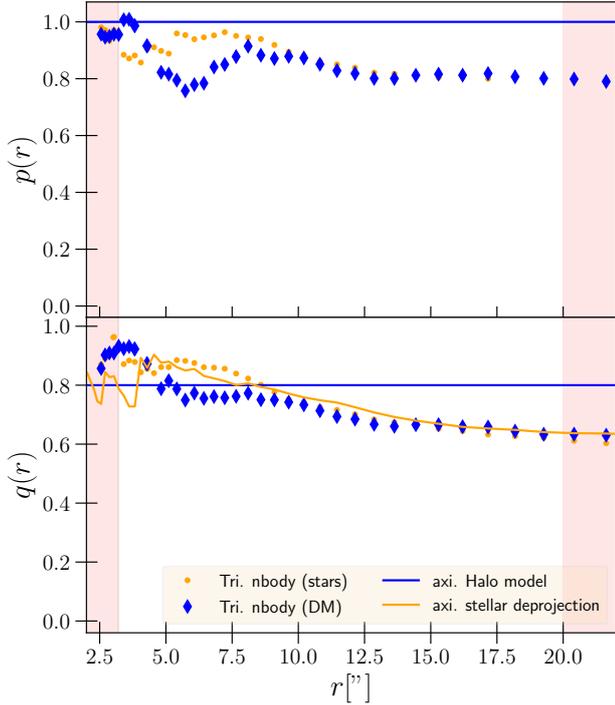


Figure 13. The shape of the N-body simulation’s density distribution approximated by an ellipsoid. *Top:* $p = b/a$, *Bottom:* $q = c/a$ where a is the semi-major axis, b the semi-intermediate axis, and c the semi-minor axis of the triaxial ellipsoid. The *Orange Dots* indicate the triaxial axis ratios of the stellar particle distribution, the *Orange Line* shows the corresponding *axisymmetric* approximation, which is also the edge-on ‘pseudo’-deprojection that we used for the dynamical modeling (but scaled by a variable mass-to-light ratio of the stellar population). The *blue diamonds* show the triaxial ratios of the dark matter particle distribution, it follows the shape of the stellar particles very closely even outside the scouring core. The *blue Line* shows the spatially constant Dark matter flattening q_{DM} of the *best* axisymmetric (oblate) orbit model we found. We sampled the axisymmetric flattening within $q_{\text{DM}} \in [0.7, 0.8, 0.9, 1.0]$. Similar to Fig. 16 we mark regions smaller than the central spatial resolution and outside the FoV in *red*.

tion is ideal. We plot the triaxial axis ratios p and q versus radius for both matter components in Fig. 13.

To obtain realistic mock observations of the N-body simulation that emulates our observational setup used for the real dEs we place the simulated galaxy at the average distance of the Virgo cluster $d = 16.5$ Mpc and projected its stellar kinematics along its *intermediate* axis into a realistic Voronoi grid with a FoV of $20''$. The number of bins and their resolution are typical for one of the higher S/N observations we obtained in the real sample. The central Voronoi bins are typically $3''$ large. This means our spatial resolution is just below

the sphere of influence $r_{\text{SOI}} \sim 4''$ and the size of the scoured core $\sim 5''$. The resulting mock observations are shown on the left side of Fig. 14. These kinematic maps are illustrated by a truncated Gauss–Hermite series up to h_4 . However, the input for the dynamical models are actually the fully non-parametric descriptions of the LOSVDs as is the case for the dE modeling. The kinematic maps of the simulation show some interesting features that are usually not observed in real galaxies: i) A positive v - h_3 correlation ii) and a negative h_4 within the scoured core. This discrepancy may tell us something about the significance of black hole scouring in real galaxies and assumptions in N-body simulations, however, an investigation is beyond the scope of this paper.

Like we did with the dE sample we use the search algorithm NOMAD and probe a total of approximately 5000 orbit models using the halo parametrization we also used for the dEs (see Sec. 3). However, we only probed a single deprojection (or inclination) for the *stellar* component. As mentioned above the N-body simulation is observed along its intermediate axis, therefore this deprojection is assumed to be edge-on. The right hand side panels in Fig. 14 illustrate the projected kinematics of the single best orbit model we were able to find to represent the simulated mock observation shown on the left side of Fig. 14. As is the case for the modeling of the Virgo dEs we opted to model the entire FoV with the orbit models, i.e. differences in individual quadrants can not be reflected as easily by the axisymmetric model. Nevertheless, the orbit model is able to reproduce all important features of the mock observations well, giving confidence that the orbit setup (e.g. number of orbits) and mass model are adequate and flexible enough to emulate the mass and kinematic of the N-body simulation.

A.2. Mass recovery and decomposition

In Fig. 15 we show the ΔAIC_p -curves for each of the nuisance parameters that generate the mass model. The stellar mass-to-light ratio $\Upsilon_* = 1.0$ is recovered by the single best orbit model. The next best orbit models AIC_p tend to have slightly larger Υ_* but the scatter is in the single digit percentage range. If one was to apply the error estimation approach we employ for the dEs sample, i.e. by calculating the spread of the 25 best orbit models we estimate an uncertainty in Υ_* of 6%. This implies the stellar mass component is well decomposed by the dynamical modeling even when the dark matter is distributed similarly as the stars over some regions of the galaxy. As long as there is a detectable difference between stars and DM in some parts of the FoV (in this case between $10''$ to $20''$) the models are able to

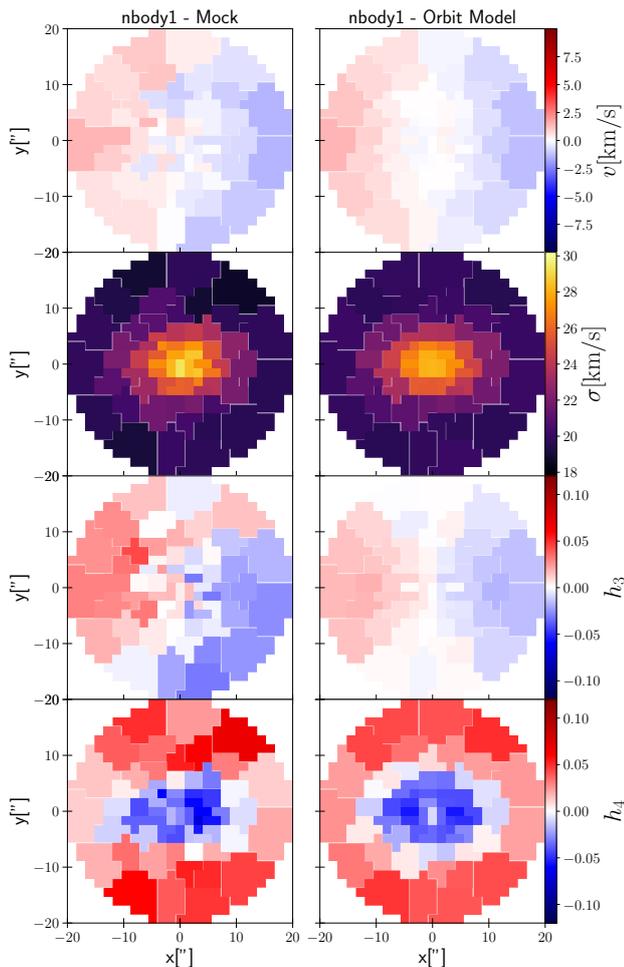


Figure 14. *Left:* The mock observations for the triaxial N-body simulation. *Right:* The corresponding best axisymmetric model we found using our orbit modeling setup. Both, mock and model LOSVDs are characterized here by a Gauss–Hermite series truncated at h_4 , higher order deviations in the actual binned LOSVDs (that are still being modelled dynamically) are not shown.

‘recognize’ that a stellar component that is simply scaled up or down is an insufficient description.

The black hole mass of the best model on the other hand is underestimated with the best model only having $0.7 \cdot 10^7$ instead of the actual SMBH mass of $1.4 \cdot 10^7$. The ΔAIC_p -curves for the black hole are very a-symmetric indicating that models with essentially no black hole and models with $M_\bullet \leq 10^7$ only show very little difference. Without prior knowledge one would likely conclude an upper limit at $M_\bullet \sim 1 \cdot 10^7$ due to the sphere of influence of the best-fitting models being close to the central resolution limit.

The parameters of the Zhao-profile suggest a scale radius $r_s \sim 0.5\text{kpc}$, i.e. the profile transitions its slope in

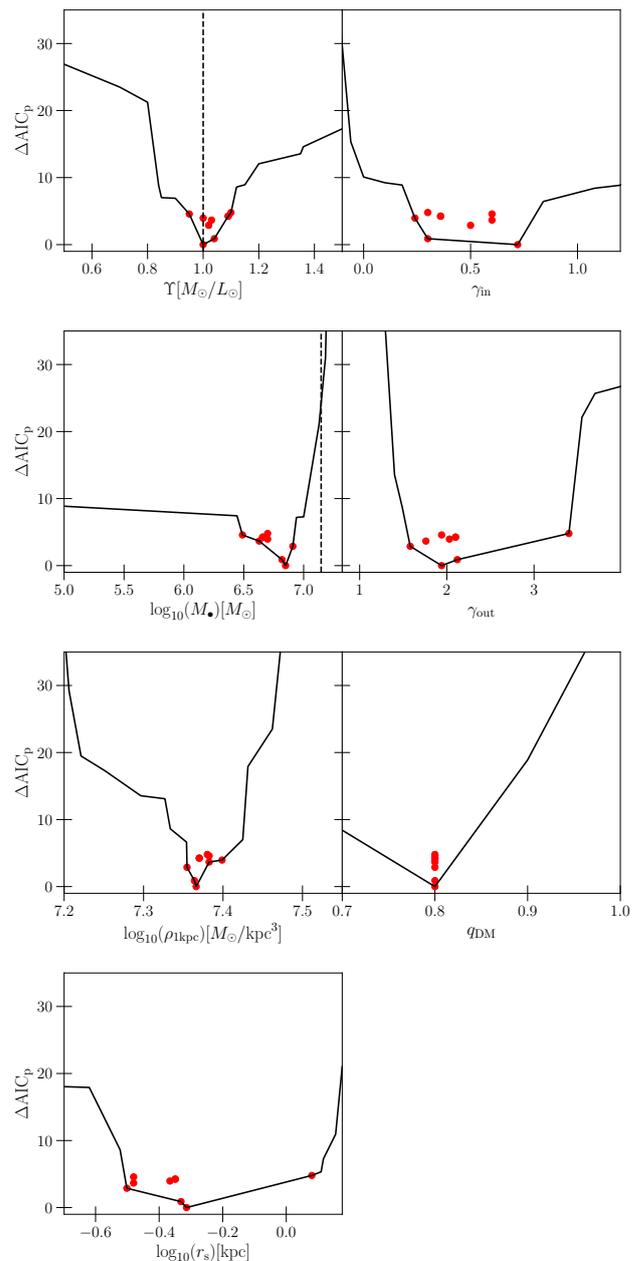


Figure 15. The minimum subtracted AIC_p -constraints for the axisymmetric orbit models of the N-body simulation. In total we calculated in the order of $\sim 5 \cdot 10^3$ orbit models. The red dots indicate the parameters of the best few models (ranked in AIC_p). The vertical dashed lines indicate the true stellar mass-to-light ratio and black hole mass.

the vicinity of the scouring core radius. However, models with a lower or larger scale radius (e.g. 0.3kpc and 1kpc) are not much worse ($\Delta\text{AIC}_p \lesssim 5$). Only for the extreme values AIC_p begins to rise more rapidly. We find similarly broad AIC_p valleys for the inner and outer slopes γ_{in} and γ_{out} of the Zhao parametrization.

Such broad AIC_p valleys in the parameters are not surprising (particularly in the more correlated parameters) and possibly even desirable if one wants to minimize biases because of erroneous halo parametrization. The dynamical models are sensitive to the actual mass distribution that is described by the above set of parameters and not the specific set of parameters or the halo parametrization itself. Therefore one should investigate how well the actual mass distributions are recovered to gauge how well the recovery worked. This is illustrated in Fig. 16 where we plot the recovered mass and anisotropy profiles of the best axisymmetric orbit model and the actual N-body simulation. We color radial regions that are within our spatial resolution or outside the FoV in red to highlight the regions where we expect the data do *not* impose strong constraints. We also plot the next few best AIC_p model we found as dashed lines and indicate their corresponding halo parameters in Fig. 15. Even though the nominal values of the strongly inter-correlated shape parameters ($\gamma_{in}, \gamma_{out}, r_s$) can differ significantly the actual mass distributions they describe are very similar. Again we conclude the parameters that ‘generate’ the halo density should be treated as nuisance parameters.

We find that the total enclosed mass $M_{tot}(< r)$ is the property constrained the best as it’s linked very directly to the gravitational potential Φ which itself determines all the orbits that constitute the orbit-superposition model. In other words, the model selection is, first and foremost, a predictor of the gravitational potential and thus the total mass (regardless of composition). One also sees that the next best few AIC_p models have very similar enclosed masses to the single best AIC_p model, though they are generally slightly further away from the true N-body mass than the single best AIC_p model which suggests that AIC_p is a consistent predictor of the enclosed mass.

However, when plotting $M_{tot}(< r)$ as a function of radius and comparing it to the actual enclosed mass of the simulation we also observe that the systematic deviation between the two varies with the radius it is measured at: Within the unresolved center (inner red region) and outside the FoV (outer red region) the differences between the ‘truth’ and the model predictions start to flare-up and diverge. The next best AIC_p models also seem to scatter *a-symmetrically* in these regions: Preferably underestimating $M_{tot}(< r)$ in the center and overestimating it at large radii outside the FoV (the latter was also noted in Gerhard et al. 1998 and Thomas et al. 2005). It seems the exact matter distribution far inside the spatial resolution (or central Voronoi elements) and far outside the FoV seems to be constrained less by the kinematic

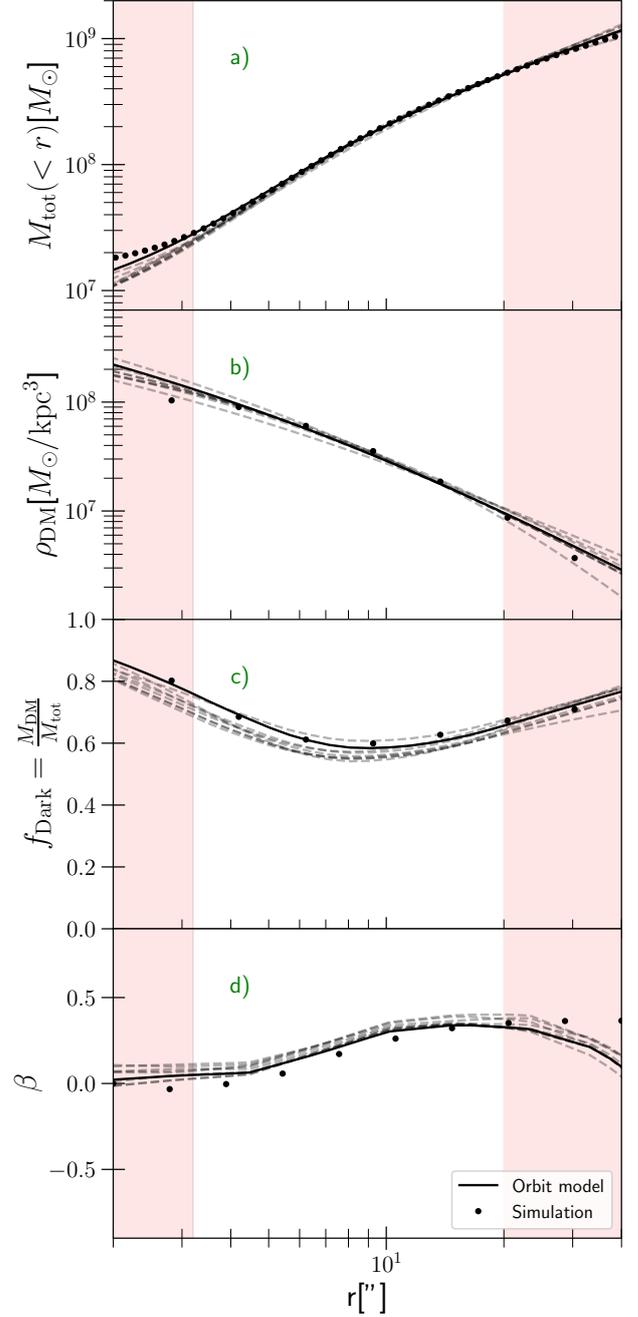


Figure 16. Mass and anisotropy structure (spherically averaged) of the simulation (*dots*) and the recovered profiles of the best *axisymmetric* orbit model we were able to find (*solid line*). The *dashed* lines show the next best models when ranking all sampled models by their AIC_p , their nuisance parameters are indicated by the red dots in Fig. 15. The *inner* red region indicates the typical size of our resolution limit. The *outer* red region marks areas outside the FoV. Panel a) the total enclosed mass within radius r , b) the local DM density, c) the (cumulated) DM fraction (including the black hole), d) the anisotropy β of the stars.

data, as several models can be constructed that differ significantly in these regions but produce essentially the same model LOSVDs (Fig. 14) and AIC_p .

Considering $M_{\text{tot}}(< r)$ is fairly well constrained within the radial range covered by the data and the fact that the models found the correct stellar mass-to-light ratios $\Upsilon_* = 1$ (cf. Fig. 15) it is not surprising that the recovery of the DM distribution (second panel) also performs similarly well. Even though the underlying dark matter distribution is more complex than the Zhao-model and quasi-degenerate with the stellar matter within $5''$ due to the mass-follows-light core, both, the dark matter density itself and its local slope $\frac{\partial \rho_{\text{DM}}}{\partial r}$ is recovered well within the white regions where $M_{\text{tot}}(< r)$ is constraint.

The dark matter inside the unresolved center does not become uncertain in a random fashion: Within the unresolved center it systematically scatters above the true cored profile of the simulation. This can be explained by a degeneracy of the *two* dark components of the model: The dark matter halo and the central black hole. As seen in Fig. 15 we find a sharp upper limit for the black hole mass of $\sim 1 \cdot 10^7$ but only very little difference in AIC_p for models with a lower black hole mass.

This underestimation of the black hole is complemented by an overestimation of the central DM density with the sphere of influence. Both these errors cancel each other if we examine the (enclosed) fraction f_{Dark} of the *combined* dark components (i.e. DM halo + black hole) versus the total enclosed mass (third panel in Fig. 16). Even in the very center the best orbit model traces the true dark fraction of the simulation well, suggesting that the separation of dark and luminous matter actually works well, merely the separation of the two dark components proves difficult. A model with an underestimated black hole mass can essentially achieve the same LOSVD predictions by assigning the ‘missing’ mass to its central dark matter halo. On the other hand, models with a much larger black hole than the true mass $M_{\bullet} = 1.4 \cdot 10^7 M_{\odot}$ are effectively ruled out because the halo can only add additional mass in the center but not subtract it.

We do not believe that previous black hole measurements in real galaxies were affected significantly by such a degeneracy, at least as long as their sphere of influence was resolved. This is because most evidence and models of real galaxies suggest that the dark matter halo fraction is almost 0 in the center, i.e. the dark matter contribution to $M_{\text{tot}}(< r)$ in unresolved scales is negligible. Which indicates that the black hole is the only significant mass in the resolved bins. Only in the cases where the dark matter fraction within the central

scales is significant or if there is suspicion of distinct unresolved stellar population (i.e. variable Υ_*) it might be worth to revisit existing black hole measurements as they may have been affected by a decomposition degeneracy. For example, Mehrgan et al. (2024) have found black-hole masses to reduced by 25% when allowing for central mass-to-light ratio gradients in a sample of massive early-type galaxies.

Assuming the two dark components are indeed degenerate it nonetheless surprising that the AIC_p -envelopes (Fig. 15) rapidly rise as $M_{\bullet} > 1 \cdot 10^7 M_{\odot}$, effectively ruling out the true mass $M_{\bullet} = 1.4 \cdot 10^7 M_{\odot}$ as too high. If the black hole constraints are degenerate but unbiased AIC_p should only start to rise rapidly for masses beyond $1.4 \cdot 10^7 M_{\odot}$. In LT we plan to identify and discuss possible reasons for this discrepancy in the upper limit for the black hole mass. As it turns out better but more computationally expensive model parametrizations can resolve this tension resulting in a better upper limit for the black hole mass $\sim 1.5 \cdot 10^7 M_{\odot}$.

While there are no obvious signs that the dEs are triaxial (see also VW-I) the real dE galaxy could at least to some degree be triaxial like the N-body simulation. To gauge how deviations from axisymmetry could affect the results for the dE sample we analyse the effects the axisymmetry assumption had on the modeling of the triaxial simulation. Two properties that we expect to be very affected by such an erroneous symmetry assumption are the intrinsic 3D *kinematics* of the stars and the *shape* (flattening) of the mass distribution. The former because the orbits of the model are restricted to an axisymmetric potential, thus, its orbits may not be as representative of the plethora of different orbits that are possible in a more general potential, and the latter because it is impossible for the model to emulate the non-axisymmetric shape of the mass distribution. This symmetry mismatch in the orbit structure and shape may in turn then negatively affect the recovered mass distributions as well (Thomas et al. 2007a).

We can examine the quality of the kinematic recovery using the orbital anisotropy parameter β :

$$\beta = 1 - \frac{\sigma_{\phi}^2 + \sigma_{\theta}^2}{2\sigma_r^2} \quad (\text{A1})$$

where the σ_i are the velocity dispersion in spherical coordinates (Binney & Tremaine 2008).

The spherically averaged β of the N-body simulation, together with the β of the best orbit models, is shown in the bottom panel of Fig. 16. Within the radial range covered by the data the anisotropy structure is recovered well within a few percent. Albeit the best AIC_p models appear to be biased slightly more radial than

the truth. This could be an expression of the incomplete orbit representation forced by the axisymmetry assumption. Similar to the mass recovery the β structure of the dynamical models starts to deviate significantly outside the FoV, again implying it is not important for a model's ability to fit the LOSVD data what exactly the mass/kinematic structure is in unconstrained regions outside the FoV.

The simulated tests show that the mass decomposition and distribution at or below unresolved scales can be misleading. In the N-body simulation this shows itself in the correlation of M_{\bullet} and the central dark matter excess. Density gradients (i.e. slopes) are recovered best for radii larger than the resolution and smaller than the FoV (white regions). The tests also show that we can expect to constrain the kinematic structure of the dE sample well even if they are slightly triaxial. However, as was the case with $M_{\text{tot}} (< r)$, the kinematic recovery suffers in areas where we have sparse or no data coverage.

A.3. Shape recovery

To evaluate the model's ability to recover the intrinsic shape of a non-axisymmetric mass distribution we approximate the N-body's particle distribution directly with *triaxial* ellipsoids (which is still a symmetry assumption but less restrictive than the axisymmetry) and determine its semi axis ratios $p = b/a$ and $q = c/a$. Where a, b , and c are the semi major-, intermediate- and minor axis of the ellipsoid. We compare this to the flattening q of the axisymmetric mass distributions which have $p = 1$ by definition since $b = a$ for the oblate, axisymmetric models.

We compare the triaxial flattening of the simulation and the dynamically recovered axisymmetric shapes of the orbit models in Fig. 13. The true triaxial shapes of the dark and baryonic components trace each other closely for the entire radial range with minor differences at intermediate radii just outside the scouring core ($5 - 8''$). As mentioned above the N-body simulation is viewed along its intermediate axis. As such, the flattening of the edge-on axisymmetric deprojection that represents the stellar component of the orbit models, is essentially a direct approximation of the minor axis flattening $q(r)$. However, the information about the additional flattening along the line of sight ($p \sim 0.8$) is lost due to the axisymmetry assumption.

While the dynamical models 'know' the axisymmetric flattening of the stellar distribution from the deprojection they have no a priori knowledge of neither the q nor p axis ratios of the dark matter distribution. In the past a long-standing point of discussion was whether the viewing angles of galaxies can be accurately recovered using only dynamical constraints. However, recently we have demonstrated that viewing angles, and consequently the intrinsic shapes, of the *luminous* component can in fact be accurately measured with dynamical models if one accounts for variation in model flexibility (e.g. Lipka & Thomas 2021; de Nicola et al. 2022). Since dynamical models trace the gravitational potential, there is no reason to assume that a flattening of the *dark* component should not be detectable as well.

Despite this though, the shapes of the dark matter component of the model are often simply assumed to be spherical even when the stellar component is obviously flattened. This assumption could heavily bias the entire mass recovery (in particular stellar mass-to-light ratios and shapes) as the model would attempt to compensate for the non-spherical halo by adjusting other components in order to approximate the total gravitational potential better. For this reason we equipped the Zhao halos we test in this study with a (radially constant) axis ratio q_{DM} as an additional free parameter. In the case of the N-body simulation we probed orbit models with four different axisymmetric flattening $q_{\text{DM}} \in [0.7, 0.8, 0.9, 1.0]$. As the AIC_p -curves in Fig. 15 suggest the orbit models show a strong preference towards an intermediate flattening of $q_{\text{DM}} = 0.8$, essentially ruling out a spherical halo. In Fig. 13 we plot the (constant) axisymmetric flattening $q_{\text{DM}} = 0.8$ of the best orbit model together with the true triaxial shapes of the simulation. While it's obvious that the halo model is not sophisticated enough to describe the halo's radial variation in q or its triaxiality ($p \neq 1$) one can see that a $q_{\text{DM}} = 0.8$ is the closest approximation for the average flattening in q and p within the radial range that is constrained by the data (white region). This suggests that we can place *unbiased* dynamical constraints on the average halo shapes of galaxies, even if the true halo is slightly triaxial or has variable flattening. In the future, a more sophisticated halo description may allow an even more accurate halo recovery. At present, however, this appears computationally unfeasible. Nevertheless, the simulated test implies that we can at least expect to infer whether the halos of the dE sample are spherical or show signs of flattening, even if they are triaxial.

REFERENCES

- Abadi, M. G., Navarro, J. F., Fardal, M., Babul, A., & Steinmetz, M. 2010, MNRAS, 407, 435, doi: [10.1111/j.1365-2966.2010.16912.x](https://doi.org/10.1111/j.1365-2966.2010.16912.x)
- Adams, J. J., Simon, J. D., Fabricius, M. H., et al. 2014, ApJ, 789, 63, doi: [10.1088/0004-637X/789/1/63](https://doi.org/10.1088/0004-637X/789/1/63)

- Akaike, H. 1973, *Information Theory and an Extension of the Maximum Likelihood Principle* (New York, NY: Springer New York), 199–213
- Akaike, H. 1974, *IEEE Transactions on Automatic Control*, 19, 716
- Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, *MNRAS*, 367, 1781, doi: [10.1111/j.1365-2966.2006.10094.x](https://doi.org/10.1111/j.1365-2966.2006.10094.x)
- Arjona-Galvez, E., Di Cintio, A., & Grand, R. J. J. 2024, arXiv e-prints, arXiv:2402.00929, doi: [10.48550/arXiv.2402.00929](https://doi.org/10.48550/arXiv.2402.00929)
- Ascasibar, Y., Yepes, G., Gottlöber, S., & Müller, V. 2004, *MNRAS*, 352, 1109, doi: [10.1111/j.1365-2966.2004.08005.x](https://doi.org/10.1111/j.1365-2966.2004.08005.x)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Audet, C., & Dennis, J. E. 2006, *SIAM Journal on Optimization*, 17, 188, doi: [10.1137/040603371](https://doi.org/10.1137/040603371)
- Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, *ApJ*, 705, 1099, doi: [10.1088/0004-637X/705/2/1099](https://doi.org/10.1088/0004-637X/705/2/1099)
- . 2010, *ApJ*, 724, 511, doi: [10.1088/0004-637X/724/1/511](https://doi.org/10.1088/0004-637X/724/1/511)
- Aumer, M., Burkert, A., Johansson, P. H., & Genzel, R. 2010, *ApJ*, 719, 1230, doi: [10.1088/0004-637X/719/2/1230](https://doi.org/10.1088/0004-637X/719/2/1230)
- Bar, N., Danieli, S., & Blum, K. 2022, *ApJL*, 932, L10, doi: [10.3847/2041-8213/ac70df](https://doi.org/10.3847/2041-8213/ac70df)
- Bender, R. 1988, *A&A*, 193, L7
- Bender, R., Surma, P., Doebereiner, S., Moellenhoff, C., & Madejsky, R. 1989, *A&A*, 217, 35
- Benetti, F., Lapi, A., Gandolfi, G., Salucci, P., & Danese, L. 2023, *ApJ*, 949, 65, doi: [10.3847/1538-4357/acc8ca](https://doi.org/10.3847/1538-4357/acc8ca)
- Bidaran, B., La Barbera, F., Pasquali, A., et al. 2022, *MNRAS*, 515, 4622, doi: [10.1093/mnras/stac2005](https://doi.org/10.1093/mnras/stac2005)
- Binggeli, B., Tammann, G. A., & Sandage, A. 1987, *AJ*, 94, 251, doi: [10.1086/114467](https://doi.org/10.1086/114467)
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* (Princeton University Press)
- Blumenthal, G. R., Faber, S. M., Flores, R., & Primack, J. R. 1986, *ApJ*, 301, 27, doi: [10.1086/163867](https://doi.org/10.1086/163867)
- Böhringer, H., Briel, U. G., Schwarz, R. A., et al. 1994, *Nature*, 368, 828, doi: [10.1038/368828a0](https://doi.org/10.1038/368828a0)
- Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, *ApJ*, 682, 964, doi: [10.1086/589327](https://doi.org/10.1086/589327)
- Bovy, J., Bahmanyar, A., Fritz, T. K., & Kallivayalil, N. 2016, *ApJ*, 833, 31, doi: [10.3847/1538-4357/833/1/31](https://doi.org/10.3847/1538-4357/833/1/31)
- Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, *MNRAS*, 415, L40, doi: [10.1111/j.1745-3933.2011.01074.x](https://doi.org/10.1111/j.1745-3933.2011.01074.x)
- . 2012, *MNRAS*, 422, 1203, doi: [10.1111/j.1365-2966.2012.20695.x](https://doi.org/10.1111/j.1365-2966.2012.20695.x)
- Brinckmann, T., Zavala, J., Rapetti, D., Hansen, S. H., & Vogelsberger, M. 2018, *MNRAS*, 474, 746, doi: [10.1093/mnras/stx2782](https://doi.org/10.1093/mnras/stx2782)
- Burkert, A. 2015, *ApJ*, 808, 158, doi: [10.1088/0004-637X/808/2/158](https://doi.org/10.1088/0004-637X/808/2/158)
- Burnham, K., & Anderson, D. 2002, *Model selection and multimodel inference: a practical information-theoretic approach* (Springer Verlag)
- Bustamante-Rosell, M. J., Noyola, E., Gebhardt, K., et al. 2021, arXiv e-prints, arXiv:2111.04770, <https://arxiv.org/abs/2111.04770>
- Cappellari, M. 2008, *MNRAS*, 390, 71, doi: [10.1111/j.1365-2966.2008.13754.x](https://doi.org/10.1111/j.1365-2966.2008.13754.x)
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, 342, 345, doi: [10.1046/j.1365-8711.2003.06541.x](https://doi.org/10.1046/j.1365-8711.2003.06541.x)
- Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, 413, 813, doi: [10.1111/j.1365-2966.2010.18174.x](https://doi.org/10.1111/j.1365-2966.2010.18174.x)
- Cappellari, M., Scott, N., Alatalo, K., et al. 2013a, *MNRAS*, 432, 1709, doi: [10.1093/mnras/stt562](https://doi.org/10.1093/mnras/stt562)
- Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2013b, *MNRAS*, 432, 1862, doi: [10.1093/mnras/stt644](https://doi.org/10.1093/mnras/stt644)
- Cataldi, P., Pedrosa, S. E., Tissera, P. B., et al. 2023, *MNRAS*, 523, 1919, doi: [10.1093/mnras/stad1601](https://doi.org/10.1093/mnras/stad1601)
- Ceverino, D., Dekel, A., & Bournaud, F. 2010, *MNRAS*, 404, 2151, doi: [10.1111/j.1365-2966.2010.16433.x](https://doi.org/10.1111/j.1365-2966.2010.16433.x)
- Chilingarian, I. V. 2009, *Monthly Notices of the Royal Astronomical Society*, 394, 1229, doi: [10.1111/j.1365-2966.2009.14450.x](https://doi.org/10.1111/j.1365-2966.2009.14450.x)
- Chisari, N. E., Koukoufilippas, N., Jindal, A., et al. 2017, *MNRAS*, 472, 1163, doi: [10.1093/mnras/stx1998](https://doi.org/10.1093/mnras/stx1998)
- Choque-Challapa, N., Smith, R., Candlish, G., Peletier, R., & Shin, J. 2019, *MNRAS*, 490, 3654, doi: [10.1093/mnras/stz2829](https://doi.org/10.1093/mnras/stz2829)
- Chua, K. T. E., Pillepich, A., Vogelsberger, M., & Hernquist, L. 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 476, doi: [10.1093/mnras/sty3531](https://doi.org/10.1093/mnras/sty3531)
- Chua, K. T. E., Vogelsberger, M., Pillepich, A., & Hernquist, L. 2022, *MNRAS*, 515, 2681, doi: [10.1093/mnras/stac1897](https://doi.org/10.1093/mnras/stac1897)
- Codis, S., Pichon, C., & Pogosyan, D. 2015, *MNRAS*, 452, 3369, doi: [10.1093/mnras/stv1570](https://doi.org/10.1093/mnras/stv1570)
- Cole, D. R., Dehnen, W., & Wilkinson, M. I. 2011, *MNRAS*, 416, 1118, doi: [10.1111/j.1365-2966.2011.19110.x](https://doi.org/10.1111/j.1365-2966.2011.19110.x)
- Comerón, S., Trujillo, I., Cappellari, M., et al. 2023, *The massive relic galaxy NGC 1277 is dark matter deficient. From dynamical models of integral-field stellar kinematics out to five effective radii.* <https://arxiv.org/abs/2303.11360>

- Conselice, C. J., Gallagher, John S., I., & Wyse, R. F. G. 2001, *ApJ*, 559, 791, doi: [10.1086/322373](https://doi.org/10.1086/322373)
- Côté, S., Carignan, C., & Freeman, K. C. 2000, *AJ*, 120, 3027, doi: [10.1086/316883](https://doi.org/10.1086/316883)
- Croft, R. A. C., Weinberg, D. H., Bolte, M., et al. 2002, *ApJ*, 581, 20, doi: [10.1086/344099](https://doi.org/10.1086/344099)
- Şen, Ş., Peletier, R. F., Boselli, A., et al. 2018, *MNRAS*, 475, 3453, doi: [10.1093/mnras/stx3254](https://doi.org/10.1093/mnras/stx3254)
- Danieli, S., van Dokkum, P., Conroy, C., Abraham, R., & Romanowsky, A. J. 2019, *ApJL*, 874, L12, doi: [10.3847/2041-8213/ab0e8c](https://doi.org/10.3847/2041-8213/ab0e8c)
- de Blok, W. J. G. 2010, *Advances in Astronomy*, 2010, 789293, doi: [10.1155/2010/789293](https://doi.org/10.1155/2010/789293)
- de Blok, W. J. G., & Bosma, A. 2002, *A&A*, 385, 816, doi: [10.1051/0004-6361:20020080](https://doi.org/10.1051/0004-6361:20020080)
- de Blok, W. J. G., Walter, F., Brinks, E., et al. 2008, *AJ*, 136, 2648, doi: [10.1088/0004-6256/136/6/2648](https://doi.org/10.1088/0004-6256/136/6/2648)
- de Graaff, A., Pillepich, A., & Rix, H.-W. 2024, *arXiv e-prints*, arXiv:2403.00907, <https://arxiv.org/abs/2403.00907>
- de Graaff, A., Rix, H.-W., Carniani, S., et al. 2023, *arXiv e-prints*, arXiv:2308.09742, doi: [10.48550/arXiv.2308.09742](https://doi.org/10.48550/arXiv.2308.09742)
- De Leo, M., Read, J. I., Noel, N. E. D., et al. 2023, *arXiv e-prints*, arXiv:2303.08838, doi: [10.48550/arXiv.2303.08838](https://doi.org/10.48550/arXiv.2303.08838)
- de Nicola, S., Neureiter, B., Thomas, J., Saglia, R. P., & Bender, R. 2022, *MNRAS*, 517, 3445, doi: [10.1093/mnras/stac2852](https://doi.org/10.1093/mnras/stac2852)
- de Souza, R. S., Rodrigues, L. F. S., Ishida, E. E. O., & Opher, R. 2011, *MNRAS*, 415, 2969, doi: [10.1111/j.1365-2966.2011.18916.x](https://doi.org/10.1111/j.1365-2966.2011.18916.x)
- Debbastista, V. P., Moore, B., Quinn, T., et al. 2008, *ApJ*, 681, 1076, doi: [10.1086/587977](https://doi.org/10.1086/587977)
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, 368, 2, doi: [10.1111/j.1365-2966.2006.10145.x](https://doi.org/10.1111/j.1365-2966.2006.10145.x)
- Dekel, A., & Silk, J. 1986, *ApJ*, 303, 39, doi: [10.1086/164050](https://doi.org/10.1086/164050)
- Del Popolo, A. 2009, *ApJ*, 698, 2093, doi: [10.1088/0004-637X/698/2/2093](https://doi.org/10.1088/0004-637X/698/2/2093)
- . 2012, *MNRAS*, 419, 971, doi: [10.1111/j.1365-2966.2011.19754.x](https://doi.org/10.1111/j.1365-2966.2011.19754.x)
- Del Popolo, A., & Le Delliou, M. 2021, *Galaxies*, 9, 123, doi: [10.3390/galaxies9040123](https://doi.org/10.3390/galaxies9040123)
- Derkenne, C., McDermid, R. M., Poci, A., et al. 2023, *MNRAS*, 522, 3602, doi: [10.1093/mnras/stad1079](https://doi.org/10.1093/mnras/stad1079)
- Diemand, J., & Moore, B. 2011, *Advanced Science Letters*, 4, 297, doi: [10.1166/asl.2011.1211](https://doi.org/10.1166/asl.2011.1211)
- Diemand, J., Moore, B., & Stadel, J. 2004, *MNRAS*, 353, 624, doi: [10.1111/j.1365-2966.2004.08094.x](https://doi.org/10.1111/j.1365-2966.2004.08094.x)
- Donato, F., Gentile, G., Salucci, P., et al. 2009, *MNRAS*, 397, 1169, doi: [10.1111/j.1365-2966.2009.15004.x](https://doi.org/10.1111/j.1365-2966.2009.15004.x)
- Dressler, A. 1980, *ApJ*, 236, 351, doi: [10.1086/157753](https://doi.org/10.1086/157753)
- Dutta Chowdhury, D., van den Bosch, F. C., van Dokkum, P., et al. 2023, *ApJ*, 949, 68, doi: [10.3847/1538-4357/acc73d](https://doi.org/10.3847/1538-4357/acc73d)
- Dutton, A. A., & Treu, T. 2014, *MNRAS*, 438, 3594, doi: [10.1093/mnras/stt2489](https://doi.org/10.1093/mnras/stt2489)
- Eftekhari, F. S., Peletier, R. F., Scott, N., et al. 2022, *MNRAS*, 517, 4714, doi: [10.1093/mnras/stac2606](https://doi.org/10.1093/mnras/stac2606)
- El-Zant, A., Shlosman, I., & Hoffman, Y. 2001, *ApJ*, 560, 636, doi: [10.1086/322516](https://doi.org/10.1086/322516)
- Elbert, O. D., Bullock, J. S., Garrison-Kimmel, S., et al. 2015, *MNRAS*, 453, 29, doi: [10.1093/mnras/stv1470](https://doi.org/10.1093/mnras/stv1470)
- Fabricius, M. H., Barnes, S., Bender, R., et al. 2008, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean & M. M. Casali, 701473, doi: [10.1117/12.787204](https://doi.org/10.1117/12.787204)
- Fabricius, M. H., Grupp, F., Bender, R., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84465K, doi: [10.1117/12.925177](https://doi.org/10.1117/12.925177)
- Fall, S. M. 1983, in *Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula, Vol. 100, 391–398
- Fall, S. M., & Efstathiou, G. 1980, *MNRAS*, 193, 189, doi: [10.1093/mnras/193.2.189](https://doi.org/10.1093/mnras/193.2.189)
- Ferrarese, L., & Merritt, D. 2000, *ApJL*, 539, L9, doi: [10.1086/312838](https://doi.org/10.1086/312838)
- Ferrarese, L., Côté, P., Jordán, A., et al. 2006, *ApJS*, 164, 334, doi: [10.1086/501350](https://doi.org/10.1086/501350)
- Frenk, C. S., White, S. D. M., Davis, M., & Efstathiou, G. 1988, *ApJ*, 327, 507, doi: [10.1086/166213](https://doi.org/10.1086/166213)
- Fukushige, T., & Makino, J. 2001, *ApJ*, 557, 533, doi: [10.1086/321666](https://doi.org/10.1086/321666)
- Gao, L., Navarro, J. F., Cole, S., et al. 2008, *MNRAS*, 387, 536, doi: [10.1111/j.1365-2966.2008.13277.x](https://doi.org/10.1111/j.1365-2966.2008.13277.x)
- Gao, L., Springel, V., & White, S. D. M. 2005, *MNRAS*, 363, L66, doi: [10.1111/j.1745-3933.2005.00084.x](https://doi.org/10.1111/j.1745-3933.2005.00084.x)
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, *ApJL*, 539, L13, doi: [10.1086/312840](https://doi.org/10.1086/312840)
- Geha, M., Guhathakurta, P., & van der Marel, R. P. 2003, *AJ*, 126, 1794, doi: [10.1086/377624](https://doi.org/10.1086/377624)
- Genina, A., Benítez-Llambay, A., Frenk, C. S., et al. 2018, *MNRAS*, 474, 1398, doi: [10.1093/mnras/stx2855](https://doi.org/10.1093/mnras/stx2855)
- Gentile, G., Salucci, P., Klein, U., Vergani, D., & Kalberla, P. 2004, *MNRAS*, 351, 903, doi: [10.1111/j.1365-2966.2004.07836.x](https://doi.org/10.1111/j.1365-2966.2004.07836.x)

- Gerhard, O., Jeske, G., Saglia, R. P., & Bender, R. 1998, *Monthly Notices of the Royal Astronomical Society*, 295, 197, doi: <https://doi.org/10.1046/j.1365-8711.1998.29511341.x>
- Gerhard, O., Kronawitter, A., Saglia, R. P., & Bender, R. 2001, *AJ*, 121, 1936, doi: [10.1086/319940](https://doi.org/10.1086/319940)
- Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A., & Nagai, D. 2004, *ApJ*, 616, 16, doi: [10.1086/424914](https://doi.org/10.1086/424914)
- Gnedin, O. Y., & Zhao, H. 2002, *MNRAS*, 333, 299, doi: [10.1046/j.1365-8711.2002.05361.x](https://doi.org/10.1046/j.1365-8711.2002.05361.x)
- Governato, F., Brook, C., Mayer, L., et al. 2010, *Nature*, 463, 203, doi: [10.1038/nature08640](https://doi.org/10.1038/nature08640)
- Governato, F., Zolotov, A., Pontzen, A., et al. 2012, *MNRAS*, 422, 1231, doi: [10.1111/j.1365-2966.2012.20696.x](https://doi.org/10.1111/j.1365-2966.2012.20696.x)
- Gunn, J. E., & Gott, J. Richard, I. 1972, *ApJ*, 176, 1, doi: [10.1086/151605](https://doi.org/10.1086/151605)
- Harko, T. 2011, *JCAP*, 2011, 022, doi: [10.1088/1475-7516/2011/05/022](https://doi.org/10.1088/1475-7516/2011/05/022)
- Hayashi, E., Navarro, J. F., & Springel, V. 2007, *MNRAS*, 377, 50, doi: [10.1111/j.1365-2966.2007.11599.x](https://doi.org/10.1111/j.1365-2966.2007.11599.x)
- Hayashi, K., Chiba, M., & Ishiyama, T. 2020, *ApJ*, 904, 45, doi: [10.3847/1538-4357/abbe0a](https://doi.org/10.3847/1538-4357/abbe0a)
- Hu, J. 2008, *MNRAS*, 386, 2242, doi: [10.1111/j.1365-2966.2008.13195.x](https://doi.org/10.1111/j.1365-2966.2008.13195.x)
- Ibata, R., Lewis, G. F., Irwin, M., Totten, E., & Quinn, T. 2001, *ApJ*, 551, 294, doi: [10.1086/320060](https://doi.org/10.1086/320060)
- Immeli, A., Samland, M., Westera, P., & Gerhard, O. 2004, *ApJ*, 611, 20, doi: [10.1086/422179](https://doi.org/10.1086/422179)
- Inoue, S., & Saitoh, T. R. 2011, *MNRAS*, 418, 2527, doi: [10.1111/j.1365-2966.2011.19873.x](https://doi.org/10.1111/j.1365-2966.2011.19873.x)
- Janz, J., Penny, S. J., Graham, A. W., Forbes, D. A., & Davies, R. L. 2017, *MNRAS*, 468, 2850, doi: [10.1093/mnras/stx634](https://doi.org/10.1093/mnras/stx634)
- Jardel, J. R., & Gebhardt, K. 2012, *ApJ*, 746, 89, doi: [10.1088/0004-637X/746/1/89](https://doi.org/10.1088/0004-637X/746/1/89)
- . 2013, *ApJL*, 775, L30, doi: [10.1088/2041-8205/775/1/L30](https://doi.org/10.1088/2041-8205/775/1/L30)
- Jardel, J. R., Gebhardt, K., Fabricius, M. H., Drory, N., & Williams, M. J. 2013, *ApJ*, 763, 91, doi: [10.1088/0004-637X/763/2/91](https://doi.org/10.1088/0004-637X/763/2/91)
- Jorgensen, I., Franx, M., & Kjaergaard, P. 1995, *MNRAS*, 273, 1097, doi: [10.1093/mnras/273.4.1097](https://doi.org/10.1093/mnras/273.4.1097)
- Katz, N., & Gunn, J. E. 1991, *ApJ*, 377, 365, doi: [10.1086/170367](https://doi.org/10.1086/170367)
- Kim, J.-h., & Lee, J. 2013, *MNRAS*, 432, 1701, doi: [10.1093/mnras/stt632](https://doi.org/10.1093/mnras/stt632)
- Klypin, A., Kravtsov, A. V., Bullock, J. S., & Primack, J. R. 2001, *ApJ*, 554, 903, doi: [10.1086/321400](https://doi.org/10.1086/321400)
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, *ApJ*, 522, 82, doi: [10.1086/307643](https://doi.org/10.1086/307643)
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102, doi: [10.1088/0004-637X/740/2/102](https://doi.org/10.1088/0004-637X/740/2/102)
- Koopmans, L. V. E., Bolton, A., Treu, T., et al. 2009, *ApJL*, 703, L51, doi: [10.1088/0004-637X/703/1/L51](https://doi.org/10.1088/0004-637X/703/1/L51)
- Kormendy, J. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 182, *Galaxy Dynamics - A Rutgers Symposium*, ed. D. R. Merritt, M. Valluri, & J. A. Sellwood, 124
- Kormendy, J., & Bender, R. 1996, *ApJL*, 464, L119, doi: [10.1086/310095](https://doi.org/10.1086/310095)
- . 2009, *ApJL*, 691, L142, doi: [10.1088/0004-637X/691/2/L142](https://doi.org/10.1088/0004-637X/691/2/L142)
- . 2013, *ApJL*, 769, L5, doi: [10.1088/2041-8205/769/1/L5](https://doi.org/10.1088/2041-8205/769/1/L5)
- Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, *ApJS*, 182, 216, doi: [10.1088/0067-0049/182/1/216](https://doi.org/10.1088/0067-0049/182/1/216)
- Kormendy, J., & Freeman, K. C. 2016, *ApJ*, 817, 84, doi: [10.3847/0004-637X/817/2/84](https://doi.org/10.3847/0004-637X/817/2/84)
- Koudmani, S., Sijacki, D., & Smith, M. C. 2022, *MNRAS*, 516, 2112, doi: [10.1093/mnras/stac2252](https://doi.org/10.1093/mnras/stac2252)
- La Barbera, F., de Carvalho, R. R., de La Rosa, I. G., et al. 2010, *MNRAS*, 408, 1313, doi: [10.1111/j.1365-2966.2010.16850.x](https://doi.org/10.1111/j.1365-2966.2010.16850.x)
- Laporte, C. F. P., & Penarrubia, J. 2015, *MNRAS*, 449, L90, doi: [10.1093/mnras/ltv008](https://doi.org/10.1093/mnras/ltv008)
- Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, *ApJ*, 237, 692, doi: [10.1086/157917](https://doi.org/10.1086/157917)
- Law, D. R., & Majewski, S. R. 2010, *ApJ*, 714, 229, doi: [10.1088/0004-637X/714/1/229](https://doi.org/10.1088/0004-637X/714/1/229)
- Le Digabel, S. 2011, *ACM Trans. Math. Softw.*, 37, doi: [10.1145/1916461.1916468](https://doi.org/10.1145/1916461.1916468)
- Lelli, F., McGaugh, S. S., & Schombert, J. M. 2016, *AJ*, 152, 157, doi: [10.3847/0004-6256/152/6/157](https://doi.org/10.3847/0004-6256/152/6/157)
- Li, P., McGaugh, S. S., Lelli, F., Schombert, J. M., & Pawlowski, M. S. 2022, *A&A*, 665, A143, doi: [10.1051/0004-6361/202243916](https://doi.org/10.1051/0004-6361/202243916)
- Li, R., Li, H., Shao, S., et al. 2019, *MNRAS*, 490, 2124, doi: [10.1093/mnras/stz2565](https://doi.org/10.1093/mnras/stz2565)
- Lin, D. N. C., & Faber, S. M. 1983, *ApJL*, 266, L21, doi: [10.1086/183971](https://doi.org/10.1086/183971)
- Lipka, M., & Thomas, J. 2021, *MNRAS*, 504, 4599, doi: [10.1093/mnras/stab1092](https://doi.org/10.1093/mnras/stab1092)
- López, P., Merchán, M. E., & Paz, D. J. 2019, *MNRAS*, 485, 5244, doi: [10.1093/mnras/stz762](https://doi.org/10.1093/mnras/stz762)
- Lovell, M. R., Pillepich, A., Genel, S., et al. 2018, *MNRAS*, 481, 1950, doi: [10.1093/mnras/sty2339](https://doi.org/10.1093/mnras/sty2339)
- Madau, P., Shen, S., & Governato, F. 2014, *ApJL*, 789, L17, doi: [10.1088/2041-8205/789/1/L17](https://doi.org/10.1088/2041-8205/789/1/L17)

- Mannerkoski, M., Johansson, P. H., Rantala, A., Naab, T., & Liao, S. 2021, *ApJL*, 912, L20, doi: [10.3847/2041-8213/abf9a5](https://doi.org/10.3847/2041-8213/abf9a5)
- Marsh, D. J. E., & Silk, J. 2014, *MNRAS*, 437, 2652, doi: [10.1093/mnras/stt2079](https://doi.org/10.1093/mnras/stt2079)
- Mashchenko, S., Couchman, H. M. P., & Wadsley, J. 2006, *Nature*, 442, 539, doi: [10.1038/nature04944](https://doi.org/10.1038/nature04944)
- McGaugh, S. S., & de Blok, W. J. G. 1998, *ApJ*, 499, 41, doi: [10.1086/305612](https://doi.org/10.1086/305612)
- Mehlert, D., Saglia, R. P., Bender, R., & Wegner, G. 2000, *A&AS*, 141, 449, doi: [10.1051/aas:2000322](https://doi.org/10.1051/aas:2000322)
- Mehrgan, K., Thomas, J., Saglia, R., et al. 2019, *ApJ*, 887, 195, doi: [10.3847/1538-4357/ab5856](https://doi.org/10.3847/1538-4357/ab5856)
- . 2024, *ApJ*, 961, 127, doi: [10.3847/1538-4357/acfe09](https://doi.org/10.3847/1538-4357/acfe09)
- Merrifield, M. R., & Kent, S. M. 1990, *AJ*, 99, 1548, doi: [10.1086/115438](https://doi.org/10.1086/115438)
- Mezcua, M., Suh, H., & Civano, F. 2019, *MNRAS*, 488, 685, doi: [10.1093/mnras/stz1760](https://doi.org/10.1093/mnras/stz1760)
- Moore, B. 1994, *Nature*, 370, 629, doi: [10.1038/370629a0](https://doi.org/10.1038/370629a0)
- Moore, B., Calcáneo-Roldán, C., Stadel, J., et al. 2001, *PhRvD*, 64, 063508, doi: [10.1103/PhysRevD.64.063508](https://doi.org/10.1103/PhysRevD.64.063508)
- Moore, B., Governato, F., Quinn, T., Stadel, J., & Lake, G. 1998, *ApJL*, 499, L5, doi: [10.1086/311333](https://doi.org/10.1086/311333)
- Mukherjee, S., Koopmans, L. V. E., Tortora, C., et al. 2022, *MNRAS*, 509, 1245, doi: [10.1093/mnras/stab3014](https://doi.org/10.1093/mnras/stab3014)
- Navarro, J. F., Eke, V. R., & Frenk, C. S. 1996a, *MNRAS*, 283, L72, doi: [10.1093/mnras/283.3.L72](https://doi.org/10.1093/mnras/283.3.L72)
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996b, *ApJ*, 462, 563, doi: [10.1086/177173](https://doi.org/10.1086/177173)
- . 1997, *ApJ*, 490, 493, doi: [10.1086/304888](https://doi.org/10.1086/304888)
- Navarro, J. F., Ludlow, A., Springel, V., et al. 2010, *MNRAS*, 402, 21, doi: [10.1111/j.1365-2966.2009.15878.x](https://doi.org/10.1111/j.1365-2966.2009.15878.x)
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 475, 624, doi: [10.1093/mnras/stx3040](https://doi.org/10.1093/mnras/stx3040)
- Nipoti, C., & Binney, J. 2015, *MNRAS*, 446, 1820, doi: [10.1093/mnras/stu2217](https://doi.org/10.1093/mnras/stu2217)
- Oh, S.-H., Brook, C., Governato, F., et al. 2011a, *AJ*, 142, 24, doi: [10.1088/0004-6256/142/1/24](https://doi.org/10.1088/0004-6256/142/1/24)
- Oh, S.-H., de Blok, W. J. G., Brinks, E., Walter, F., & Kennicutt, Robert C., J. 2011b, *AJ*, 141, 193, doi: [10.1088/0004-6256/141/6/193](https://doi.org/10.1088/0004-6256/141/6/193)
- Oh, S.-H., Hunter, D. A., Brinks, E., et al. 2015, *AJ*, 149, 180, doi: [10.1088/0004-6256/149/6/180](https://doi.org/10.1088/0004-6256/149/6/180)
- Orkney, M. D. A., Taylor, E., Read, J. I., et al. 2023, arXiv e-prints, arXiv:2302.12818, doi: [10.48550/arXiv.2302.12818](https://doi.org/10.48550/arXiv.2302.12818)
- Orkney, M. D. A., Read, J. I., Rey, M. P., et al. 2021, *MNRAS*, 504, 3509, doi: [10.1093/mnras/stab1066](https://doi.org/10.1093/mnras/stab1066)
- Pacucci, F., Ni, Y., & Loeb, A. 2023, Extreme Tidal Stripping May Explain the Overmassive Black Hole in Leo I: a Proof of Concept. <https://arxiv.org/abs/2309.02487>
- Partmann, C., Naab, T., Rantala, A., et al. 2023, arXiv e-prints, arXiv:2310.08079, doi: [10.48550/arXiv.2310.08079](https://doi.org/10.48550/arXiv.2310.08079)
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *A&A*, 412, 45, doi: [10.1051/0004-6361:20031411](https://doi.org/10.1051/0004-6361:20031411)
- Paudel, S., Yoon, S.-J., Yoo, J., et al. 2023, *ApJS*, 265, 57, doi: [10.3847/1538-4365/acbfa7](https://doi.org/10.3847/1538-4365/acbfa7)
- Peebles, P. J. E. 1969, *ApJ*, 155, 393, doi: [10.1086/149876](https://doi.org/10.1086/149876)
- Peter, A. H. G., Rocha, M., Bullock, J. S., & Kaplinghat, M. 2013, *MNRAS*, 430, 105, doi: [10.1093/mnras/sts535](https://doi.org/10.1093/mnras/sts535)
- Peters, S. P. C., van der Kruit, P. C., Allen, R. J., & Freeman, K. C. 2017, *MNRAS*, 464, 65, doi: [10.1093/mnras/stw2101](https://doi.org/10.1093/mnras/stw2101)
- Plana, H., Amram, P., Mendes de Oliveira, C., & Balkowski, C. 2010, *AJ*, 139, 1, doi: [10.1088/0004-6256/139/1/1](https://doi.org/10.1088/0004-6256/139/1/1)
- Poci, A., Cappellari, M., & McDermid, R. M. 2017, *MNRAS*, 467, 1397, doi: [10.1093/mnras/stx101](https://doi.org/10.1093/mnras/stx101)
- Pontzen, A., & Governato, F. 2012, *MNRAS*, 421, 3464, doi: [10.1111/j.1365-2966.2012.20571.x](https://doi.org/10.1111/j.1365-2966.2012.20571.x)
- . 2014, *Nature*, 506, 171, doi: [10.1038/nature12953](https://doi.org/10.1038/nature12953)
- Porciani, C., Dekel, A., & Hoffman, Y. 2002, *MNRAS*, 332, 325, doi: [10.1046/j.1365-8711.2002.05305.x](https://doi.org/10.1046/j.1365-8711.2002.05305.x)
- Posacki, S., Cappellari, M., Treu, T., Pellegrini, S., & Ciotti, L. 2014, *Monthly Notices of the Royal Astronomical Society*, 446, 493, doi: [10.1093/mnras/stu2098](https://doi.org/10.1093/mnras/stu2098)
- Posti, L., & Helmi, A. 2019, *A&A*, 621, A56, doi: [10.1051/0004-6361/201833355](https://doi.org/10.1051/0004-6361/201833355)
- Pulsoni, C., Gerhard, O., Arnaboldi, M., et al. 2021, *A&A*, 647, A95, doi: [10.1051/0004-6361/202039166](https://doi.org/10.1051/0004-6361/202039166)
- Quilis, V., Moore, B., & Bower, R. 2000, *Science*, 288, 1617, doi: [10.1126/science.288.5471.1617](https://doi.org/10.1126/science.288.5471.1617)
- Rantala, A., Johansson, P. H., Naab, T., Thomas, J., & Frigo, M. 2018, *ApJ*, 864, 113, doi: [10.3847/1538-4357/aada47](https://doi.org/10.3847/1538-4357/aada47)
- Rantala, A., Pihajoki, P., Johansson, P. H., et al. 2017, *ApJ*, 840, 53, doi: [10.3847/1538-4357/aa6d65](https://doi.org/10.3847/1538-4357/aa6d65)
- Read, J. I., Agertz, O., & Collins, M. L. M. 2016, *MNRAS*, 459, 2573, doi: [10.1093/mnras/stw713](https://doi.org/10.1093/mnras/stw713)
- Read, J. I., & Gilmore, G. 2005, *MNRAS*, 356, 107, doi: [10.1111/j.1365-2966.2004.08424.x](https://doi.org/10.1111/j.1365-2966.2004.08424.x)
- Remus, R.-S., Burkert, A., Dolag, K., et al. 2013, *ApJ*, 766, 71, doi: [10.1088/0004-637X/766/2/71](https://doi.org/10.1088/0004-637X/766/2/71)
- Remus, R.-S., Dolag, K., Naab, T., et al. 2017, *MNRAS*, 464, 3742, doi: [10.1093/mnras/stw2594](https://doi.org/10.1093/mnras/stw2594)

- Robles, V. H., & Matos, T. 2012, *MNRAS*, 422, 282, doi: [10.1111/j.1365-2966.2012.20603.x](https://doi.org/10.1111/j.1365-2966.2012.20603.x)
- Romanowsky, A. J., & Fall, S. M. 2012, *ApJS*, 203, 17, doi: [10.1088/0067-0049/203/2/17](https://doi.org/10.1088/0067-0049/203/2/17)
- Romero-Gómez, J., Peletier, R. F., Aguerri, J. A. L., & Smith, R. 2024, arXiv e-prints, arXiv:2404.15519, doi: [10.48550/arXiv.2404.15519](https://doi.org/10.48550/arXiv.2404.15519)
- Romero-Gómez, J., Aguerri, J. A. L., Peletier, R. F., et al. 2023a, *Monthly Notices of the Royal Astronomical Society*, 527, 9715, doi: [10.1093/mnras/stad3801](https://doi.org/10.1093/mnras/stad3801)
- Romero-Gómez, J., Peletier, R. F., Aguerri, J. A. L., et al. 2023b, *Monthly Notices of the Royal Astronomical Society*, 522, 130, doi: [10.1093/mnras/stad953](https://doi.org/10.1093/mnras/stad953)
- Rubin, V. C., & Ford, W. K. 1970, *The Astrophysical Journal*, 159, 379, <https://api.semanticscholar.org/CorpusID:122756867>
- Ryś, A., Falcón-Barroso, J., & van de Ven, G. 2013, *MNRAS*, 428, 2980, doi: [10.1093/mnras/sts245](https://doi.org/10.1093/mnras/sts245)
- Sandage, A., & Binggeli, B. 1984, *AJ*, 89, 919, doi: [10.1086/113588](https://doi.org/10.1086/113588)
- Schwarzschild, M. 1979, *ApJ*, 232, 236, doi: [10.1086/157282](https://doi.org/10.1086/157282)
- Scott, N., Eftekhari, F. S., Peletier, R. F., et al. 2020, *MNRAS*, 497, 1571, doi: [10.1093/mnras/staa2042](https://doi.org/10.1093/mnras/staa2042)
- Seo, M., & Ann, H. B. 2023, *MNRAS*, 520, 5521, doi: [10.1093/mnras/stad425](https://doi.org/10.1093/mnras/stad425)
- Sharma, G., Freundlich, J., van de Ven, G., et al. 2023a, arXiv e-prints, arXiv:2309.04541, doi: [10.48550/arXiv.2309.04541](https://doi.org/10.48550/arXiv.2309.04541)
- Sharma, R. S., Brooks, A. M., Tremmel, M., Bellovary, J., & Quinn, T. R. 2023b, *ApJ*, 957, 16, doi: [10.3847/1538-4357/ace046](https://doi.org/10.3847/1538-4357/ace046)
- Sharma, R. S., Brooks, A. M., Tremmel, M., et al. 2022, arXiv e-prints, arXiv:2203.05580, <https://arxiv.org/abs/2203.05580>
- Shen, Z., Danieli, S., van Dokkum, P., et al. 2021, *ApJL*, 914, L12, doi: [10.3847/2041-8213/ac0335](https://doi.org/10.3847/2041-8213/ac0335)
- Silk, J. 2017, *ApJL*, 839, L13, doi: [10.3847/2041-8213/aa67da](https://doi.org/10.3847/2041-8213/aa67da)
- Skillman, E. D., & Bender, R. 1995, in *Revista Mexicana de Astronomía y Astrofísica Conference Series*, Vol. 3, *Revista Mexicana de Astronomía y Astrofísica Conference Series*, ed. M. Pena & S. Kurtz, 25
- Smith, R., Fellhauer, M., & Assmann, P. 2012, *MNRAS*, 420, 1990, doi: [10.1111/j.1365-2966.2011.20077.x](https://doi.org/10.1111/j.1365-2966.2011.20077.x)
- Somerville, R. S., Behroozi, P., Pandya, V., et al. 2018, *MNRAS*, 473, 2714, doi: [10.1093/mnras/stx2040](https://doi.org/10.1093/mnras/stx2040)
- Spergel, D. N., & Steinhardt, P. J. 2000, *PhRvL*, 84, 3760, doi: [10.1103/PhysRevLett.84.3760](https://doi.org/10.1103/PhysRevLett.84.3760)
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, *ApJS*, 148, 175, doi: [10.1086/377226](https://doi.org/10.1086/377226)
- Springel, V., Frenk, C. S., & White, S. D. M. 2006, *Nature*, 440, 1137, doi: [10.1038/nature04805](https://doi.org/10.1038/nature04805)
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629, doi: [10.1038/nature03597](https://doi.org/10.1038/nature03597)
- Stadel, J., Potter, D., Moore, B., et al. 2009, *MNRAS*, 398, L21, doi: [10.1111/j.1745-3933.2009.00699.x](https://doi.org/10.1111/j.1745-3933.2009.00699.x)
- Steinhauser, D., Schindler, S., & Springel, V. 2016, *A&A*, 591, A51, doi: [10.1051/0004-6361/201527705](https://doi.org/10.1051/0004-6361/201527705)
- Sybilka, A., Lisker, T., Kuntschner, H., et al. 2017, *MNRAS*, 470, 815, doi: [10.1093/mnras/stx1138](https://doi.org/10.1093/mnras/stx1138)
- Tenneti, A., Mandelbaum, R., Di Matteo, T., Feng, Y., & Khandai, N. 2014, *MNRAS*, 441, 470, doi: [10.1093/mnras/stu586](https://doi.org/10.1093/mnras/stu586)
- Thomas, J., Jesseit, R., Naab, T., et al. 2007a, *MNRAS*, 381, 1672, doi: [10.1111/j.1365-2966.2007.12343.x](https://doi.org/10.1111/j.1365-2966.2007.12343.x)
- Thomas, J., & Lipka, M. 2022, *MNRAS*, 514, 6203, doi: [10.1093/mnras/stac1581](https://doi.org/10.1093/mnras/stac1581)
- Thomas, J., Saglia, R. P., Bender, R., Erwin, P., & Fabricius, M. 2014, *ApJ*, 782, 39, doi: [10.1088/0004-637X/782/1/39](https://doi.org/10.1088/0004-637X/782/1/39)
- Thomas, J., Saglia, R. P., Bender, R., et al. 2005, *MNRAS*, 360, 1355, doi: [10.1111/j.1365-2966.2005.09139.x](https://doi.org/10.1111/j.1365-2966.2005.09139.x)
- . 2007b, *MNRAS*, 382, 657, doi: [10.1111/j.1365-2966.2007.12434.x](https://doi.org/10.1111/j.1365-2966.2007.12434.x)
- . 2009, *ApJ*, 691, 770, doi: [10.1088/0004-637X/691/1/770](https://doi.org/10.1088/0004-637X/691/1/770)
- . 2004, *MNRAS*, 353, 391, doi: [10.1111/j.1365-2966.2004.08072.x](https://doi.org/10.1111/j.1365-2966.2004.08072.x)
- Toloba, E., Guhathakurta, P., Peletier, R. F., et al. 2014, *ApJS*, 215, 17, doi: [10.1088/0067-0049/215/2/17](https://doi.org/10.1088/0067-0049/215/2/17)
- Toloba, E., Guhathakurta, P., Boselli, A., et al. 2015, *ApJ*, 799, 172, doi: [10.1088/0004-637X/799/2/172](https://doi.org/10.1088/0004-637X/799/2/172)
- Tortora, C., La Barbera, F., & Napolitano, N. R. 2016, *MNRAS*, 455, 308, doi: [10.1093/mnras/stv2250](https://doi.org/10.1093/mnras/stv2250)
- Tortora, C., La Barbera, F., Napolitano, N. R., de Carvalho, R. R., & Romanowsky, A. J. 2012, *MNRAS*, 425, 577, doi: [10.1111/j.1365-2966.2012.21506.x](https://doi.org/10.1111/j.1365-2966.2012.21506.x)
- Tortora, C., Napolitano, N. R., Cardone, V. F., et al. 2010, *MNRAS*, 407, 144, doi: [10.1111/j.1365-2966.2010.16938.x](https://doi.org/10.1111/j.1365-2966.2010.16938.x)
- Tortora, C., Napolitano, N. R., Saglia, R. P., et al. 2014, *MNRAS*, 445, 162, doi: [10.1093/mnras/stu1712](https://doi.org/10.1093/mnras/stu1712)
- Tortora, C., Posti, L., Koopmans, L. V. E., & Napolitano, N. R. 2019, *MNRAS*, 489, 5483, doi: [10.1093/mnras/stz2320](https://doi.org/10.1093/mnras/stz2320)
- van der Marel, R. P., & Franx, M. 1993, *ApJ*, 407, 525, doi: [10.1086/172534](https://doi.org/10.1086/172534)
- van Dokkum, P., Abraham, R., Brodie, J., et al. 2016, *ApJL*, 828, L6, doi: [10.3847/2041-8205/828/1/L6](https://doi.org/10.3847/2041-8205/828/1/L6)
- van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, *Nature*, 555, 629, doi: [10.1038/nature25767](https://doi.org/10.1038/nature25767)

- van Zee, L., Skillman, E. D., & Haynes, M. P. 2004, *AJ*, 128, 121, doi: [10.1086/421368](https://doi.org/10.1086/421368)
- Vera-Ciro, C., & Helmi, A. 2013, *ApJL*, 773, L4, doi: [10.1088/2041-8205/773/1/L4](https://doi.org/10.1088/2041-8205/773/1/L4)
- Vogelsberger, M., Zavala, J., Cyr-Racine, F.-Y., et al. 2016, *MNRAS*, 460, 1399, doi: [10.1093/mnras/stw1076](https://doi.org/10.1093/mnras/stw1076)
- Walker, M. G., & Peñarrubia, J. 2011, *ApJ*, 742, 20, doi: [10.1088/0004-637X/742/1/20](https://doi.org/10.1088/0004-637X/742/1/20)
- Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, *ApJ*, 568, 52, doi: [10.1086/338765](https://doi.org/10.1086/338765)
- Wegg, C., Gerhard, O., & Bieth, M. 2019, *MNRAS*, 485, 3296, doi: [10.1093/mnras/stz572](https://doi.org/10.1093/mnras/stz572)
- Weller, E. J., Pacucci, F., Natarajan, P., & Matteo, T. D. 2023, Over-massive Central Black Holes in the Cosmological Simulations ASTRID and Illustris TNG50. <https://arxiv.org/abs/2305.02335>
- Williams, L. L. R., Babul, A., & Dalcanton, J. J. 2004, *ApJ*, 604, 18, doi: [10.1086/381722](https://doi.org/10.1086/381722)
- Zhao, H. 1996, *MNRAS*, 278, 488, doi: [10.1093/mnras/278.2.488](https://doi.org/10.1093/mnras/278.2.488)
- Zöllner, R., Kluge, M., Staiger, B., & Bender, R. 2023, arXiv e-prints, arXiv:2310.09330, doi: [10.48550/arXiv.2310.09330](https://doi.org/10.48550/arXiv.2310.09330)

Chapter 5

Conclusions

In this thesis I have set out to improve the accuracy and precision of the statistical inference with Schwarzschild models. As a first application of the new and improved modelling machinery, I addressed several science questions concerning dwarf ellipticals.

In Lipka & Thomas (2021) and Thomas & Lipka (2022) the AIC_p model selection approach was introduced and tested. I presented how Schwarzschild modelling fundamentally is a model selection problem and not a parameter estimation problem. As such, the model flexibility of different trial models can not be ignored, because more flexible models will be artificially favoured otherwise. I established a novel ansatz that allows one to measure the model flexibility of Schwarzschild models, and more generally, non-linear statistical models. Using the inclination as an example, I demonstrated how the parameter estimation paradigm has led to an edge-on bias in the modelling constraints, which can be corrected if the variability of differently inclined models is accounted for with a suitable model selection approach. I investigated which of the manifold of existing model selection criteria is most suitable for our applications, and found that an Akaike selection provides optimal results. I stress-tested the new approach on a number of simulations and the galaxy NGC 3368, successfully recovering their viewing angles. Similarly, the accuracy and precision of the mass recovery was improved. The applications of the new approach go well beyond just selecting the best candidate model. Within the Schwarzschild modelling pipeline, we now use it to optimize the recovery of LOSVDs from the spectra and the regularization of the orbital weights. Outside our own modelling pipeline, the versatility of the approach may prove to be useful for statistical modelling in general.

Using high resolution spectra obtained with the IFU spectrograph VIRUS-W, I analysed a sample of dwarf ellipticals with the improved modelling pipeline. Due to their similar photometric structure, dwarf ellipticals are commonly believed to be the remnants of transformed late-type dwarfs. In my thesis I presented the first study of dEs which uses advanced dynamical models that are able to constrain their decomposed 3D mass and orbit structure. The insights gained from the modelling has allowed a much more complete comparison of possible connections (or differences) between dwarf LTGs and dEs. The comparison suggests that dEs are not simply transformed late-type dwarfs, but, more specifically, transformed late-type dwarfs that have assembled at high redshift. The anal-

ysis of the *stellar* component of the dEs which was published in Lipka et al. (in press) shows the following:

- I showed that measuring the low velocity dispersions (20-30 km/s) of dEs accurately, requires a very high spectral resolution ($R \gtrsim 5000$). Previous studies of dEs have often overestimated the dispersions significantly due to their lower resolution. The dEs show diverse, but strong, higher-order kinematic features in their LOSVDs. Many dEs have a central h_4 peak and follow the $v - h_3$ anti-correlation known from massive galaxies.
- Apart from some very small and unresolved central nuclei, the dEs are *spatially* homogeneous. The dynamically measured mass-to-light ratio gradients are close to zero throughout the galaxy, as are gradients in the age, metallicity and abundance ratios. This suggests the main body of the dwarf galaxies consists of a single stellar population that has formed uniformly, and star-formation has stopped more or less simultaneously throughout the galaxy.
- The spatial homogeneity stands in opposition to the heterogeneity between different dEs. The analysis of their stellar populations suggest some dEs were quenched rather recently within the last few Gyrs, while others were quenched shortly after reionization. The abundance ratios suggest that many of them had complex, bursty or prolonged star formation histories.
- The initial mass function of dEs is on average consistent with a Kroupa function. Within the dE sample the IMF parameter is anti-correlated with age which can be explained by a variation of the IMF with the galaxy formation epoch and/or varying degrees of extended star formation history.
- The angular momentum of dEs is suppressed compared to more massive ETGs. This leads to a trichotomy in the sequence of ETGs. Dwarf ETGs ($\log(M_*/M_\odot) \lesssim 9.5$) and giant cored ETGs ($\log(M_*/M_\odot) \lesssim 11.5$) have small/moderate angular momentum, while the angular momentum peaks in between, at around $\log(M_*/M_\odot) \sim 11$.
- The anisotropy structure of dEs is nearly isotropic in spherical coordinates and does not change much with radius. When dEs have a flattened stellar structure, they are so because they have a higher kinetic energy (ordered + unordered motion) in the equatorial plane, but less so in the radial and z-direction. This orbit structure makes (major) mergers unlikely and may be interpreted as a relic of its discy late-type progenitor.

The analysis of the dark matter and black holes, published in Lipka et al. (in press), shows the following:

- The halos of dEs are very round and have moderate central logarithmic slopes $\in [-1.4, -0.5]$. These two observational halo constraints are in tension with dark

matter only Λ CDM simulations. For simulations that include baryonic physics, the tension becomes milder and may be resolved completely in the future by improved simulations. Therefore, there is no definite evidence that more exotic physics have to be invoked to explain the results.

- The total density slopes of baryons and dark matter are shallower ($\gamma \sim -1.5$) than those of more massive ETGs ($\gamma \sim -2.1$). Still, the total density slopes of dEs are significantly steeper than dwarf LTGs of the same stellar mass.
- The central dynamics of dEs is dominated by the distribution of baryons. The black hole masses of dEs are too small to be resolved with the VIRUS-W spectrograph, but the dynamical models provide an upper limit of $M_{\bullet} \lesssim 10^6 M_{\odot}$. Within one stellar effective radius, the dark matter fractions are small to moderate (as opposed to the fractions observed in dwarf spirals).
- The average density and central slopes of the dE halos are strongly correlated with the environment they reside in. In Virgo's centre, the halos of the dEs are dense, cuspy, and have low angular momentum. The opposite is the case for the dEs that reside in the more isolated outskirts of Virgo.
- The results suggest that the halos of dEs do not have a universal profile after gravitational assembly. Depending on the conditions during their assembly, a halo turns out to be more cored or cuspy. After assembly, internal feedback (e.g. from star-formation) and external influences (e.g. by stripping) affect the halos to a lesser degree.

All together, the results convey that today's dEs can *not* be directly transformed from today's dwarf spirals. Instead, the spiral progenitors of today's dEs have all formed at high redshift in more extreme conditions than today's dwarf spirals. In these extreme conditions, some of them were quenched shortly after assembly, while others (the more isolated and massive ones) were barely able to hold onto their gas such that they experienced multiple star-formation bursts. However, even the latter were eventually quenched, likely as they entered the denser part of the cluster.

In the mass regime of the dEs not many galaxies have been investigated with sophisticated dynamical modelling techniques. Very isolated field dEs and other morphological subtypes of quiescent and star-forming dwarfs will need to be modelled and compared to in the future. At the same time, the modelling machinery and its analysis have to be improved continuously, in parallel to the increasing capabilities of the next generation of telescopes. Only robust and unbiased dynamical models that exploit the full information that is present in the data will enable us to understand the diversity and origin of galaxies more thoroughly.

Appendix A

A simple data-driven method to optimize the penalty strengths of penalized models and its application to non-parametric smoothing

Bibliographic information

Jens Thomas, Mathias Lipka - A simple data-driven method to optimize the penalty strengths of penalized models and its application to non-parametric smoothing, Monthly Notices of the Royal Astronomical Society, Volume 514, Issue 4, pp.6203-6214
DOI: [10.1093/mnras/stac1581](https://doi.org/10.1093/mnras/stac1581)

Copyright

©2022 MNRAS
reprinted on pages 176- 187

A simple data-driven method to optimize the penalty strengths of penalized models and its application to non-parametric smoothing

Jens Thomas ^{1,2}★ and Mathias Lipka ^{1,2}

¹Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching, Germany

²Universitäts-Sternwarte München, Ludwig-Maximilians-Universität, Scheinerstrasse 1, D-81679 München, Germany

Accepted 2022 May 31. Received 2022 May 30; in original form 2022 March 11

ABSTRACT

Information of interest can often only be extracted from data by model fitting. When the functional form of such a model cannot be deduced from first principles, one has to make a choice between different possible models. A common approach in such cases is to minimize the information loss in the model by trying to reduce the number of fit variables (or the model flexibility, respectively) as much as possible while still yielding an acceptable fit to the data. Model selection via the Akaike information criterion (AIC) provides such an implementation of Occam’s razor. We argue that the same principles can be applied to optimize the penalty strength of a penalized maximum-likelihood model. However, while in typical applications AIC is used to choose from a finite, discrete set of maximum-likelihood models, the penalty optimization requires to select out of a continuum of candidate models and these models violate the maximum-likelihood condition. We derive a generalized information criterion AIC_p that encompasses this case. It naturally involves the concept of effective free parameters, which is very flexible and can be applied to any model, be it linear or non-linear, parametric or non-parametric, and with or without constraint equations on the parameters. We show that the generalized AIC_p allows an optimization of any penalty strength without the need of separate Monte Carlo simulations. As an example application, we discuss the optimization of the smoothing in non-parametric models, which has many applications in astrophysics, like in dynamical modelling, spectral fitting, or gravitational lensing.

Key words: methods: data analysis – methods: numerical – methods: statistical – galaxies: kinematics and dynamics.

1 INTRODUCTION

Very often, the information one aims to extract from a set of data points is not an observable itself. Instead, one has to infer the information by fitting a *model* to the observed data. Sometimes, when one has a clear understanding of the processes involved in generating the data, the functional form of the model can be deduced from first principles. In this case, one is only faced with the problem of finding the optimal parameters of the model, whereas the form of the model is fixed.

In other cases, however, one may not have such a clear picture that allows one to deduce the form of the model. Instead, one may have measured two quantities x and y that happen to be correlated without knowing the underlying form of the relation of x and y . Then, the first question becomes *how* this correlation can be characterized: For example, can it be described by a linear model (two parameters), a parabolic model (three parameters), a third-order polynomial (four parameters), etc.? A simple principle often followed in this task of *model selection* is Occam’s razor, which is to choose the model with the smallest number of free parameters (the ‘simplest’ one) that still describes the data well. Mathematical implementations of this principle are offered by information theory: One of them, the Akaike information criterion (AIC; Akaike 1973, 1974), is frequently used

to judge models based upon the increase of model complexity (and, hence, of information *loss*) against the improvement in the goodness of fit (cf. Section 4).

Many applications of model selection deal with categorically or structurally different models like in the order-selection problem briefly described earlier. The model selection character of such problems is evident, not least because the models that are compared are represented by *different fitting functions*. However, a model is not only characterized by its fitting function. Equally important are the model *parameters*. Are the parameter ranges unlimited or subject to equality or inequality constraints? And can the parameters vary independently or are they subject to implicit correlations? The answers to these questions can change the behaviour of a model effectively as much as a change in the fitting function can do.

Consider, for example, a third-order polynomial as briefly mentioned earlier. It has four model parameters and its model complexity is much higher than that of a straight line with just two parameters. Suppose that for some reason the fit with this third-order model is subject to a penalty. A penalty proportional to the square of the second derivative is frequently used in non-parametric fits to keep a model smooth and to prevent it from overfitting the data, i.e. fitting the noise. Minimizing the second derivative – the curvature – means to make the model a straight line in our case. A penalty like this often comes along with one or more additional parameters that allow to adjust the relative strength of the penalty over the achievable goodness of fit. In the above example, when we choose the penalty strength such that

* E-mail: jthomas@mpe.mpg.de

the weight of the penalty actually vanishes, then the model behaves like an ordinary third-order fit with four independent parameters. However, if we choose the penalty strength such that it dominates the fit, then the same fitting function will effectively behave like a linear model with only two parameters.

Classically, one considers the penalty as a modification of a *specific model* and the penalty strength as a nuisance parameter of that model. However, the above illustrates that for a model's behaviour, a change of the fitting function or a change of the penalty strength can have equivalent effects. It seems therefore promising to try and view penalized maximum-likelihood models from a slightly different perspective and to reinterpret the penalty as being a function that implicitly spawns a whole new *family of different models*. While different fitting functions will typically lead to a discrete set of candidate models, the penalty term allows to generate a continuum of models. The penalty strength is the natural parameter in this continuum to distinguish between different models. Treating the penalty strength as a parameter of a family of models transforms the task of optimizing this strength into a model selection problem. In other words, it allows to adopt model selection techniques to solve the optimization problem of the penalty strength. The main challenge when dealing with a continuum of models generated via a penalty term as described earlier is that the fitting function is constant. Thus, the number of fitted parameters is invariable and cannot serve as a measure of model complexity anymore. Hence, one needs to adapt the model selection strategy.

The goal of this paper is to elaborate the above outlined ideas in detail. We show how the classical ideas of model selection can be generalized to models that do not fulfil the maximum-likelihood condition. This includes penalized maximum-likelihood models in particular. Our generalization can be naturally formulated using an intuitive, generalized concept of free parameters. We show how the generalized model selection can be used to optimize the penalty strength of penalized models. As a specific example application, we provide a simple recipe that is based on information theory and that allows to optimize the smoothing of any model. The method is purely based on the measurement data at hand and does not require separate Monte Carlo simulations (e.g. for calibration of the smoothing). It is flexible and can be applied to linear as well as non-linear models and to models with or without constraint equations on their parameters. We will use the example of an emission line model to illustrate the method and the underlying concepts. It should be noted that the method is, however, not restricted to smoothing problems and can be applied to any penalized maximum-likelihood model.

In Section 2, we introduce the mock data set that is inspired by the problem of fitting the shape of an emission line. In Section 3, we introduce two example models intended to describe these mock data: a non-linear parametric model and a linear non-parametric model. In Section 4, we recall the basics of model selection for maximum-likelihood models without penalties. In Section 5, we introduce a bootstrap method to compute the effective number of parameters that can be applied to a large class of models. In Section 6, we show how the number of effective parameters should be used within model selection and sketch the derivation of a generalized model selection criterion for penalized maximum-likelihood models. In Section 7, we apply this generalized model selection criterion to our toy problem and show how it can be used to select the right order in the parametric approach or the optimal smoothing in the non-parametric models and how to choose between the two approaches. In Section 8, we extensively discuss the efficiency of the method. The paper ends with a summary in Section 9.

Table 1. The Gauss–Hermite coefficients of the generating (input) model. The underlying Gaussian function had $\mu = 0$, $\sigma = 350$, and $\gamma = 1$.

Order n	3	4	5	6	7	8	9	10
Value	0.0	0.1	0.05	0.1	-0.05	0.0	0.0	0.2

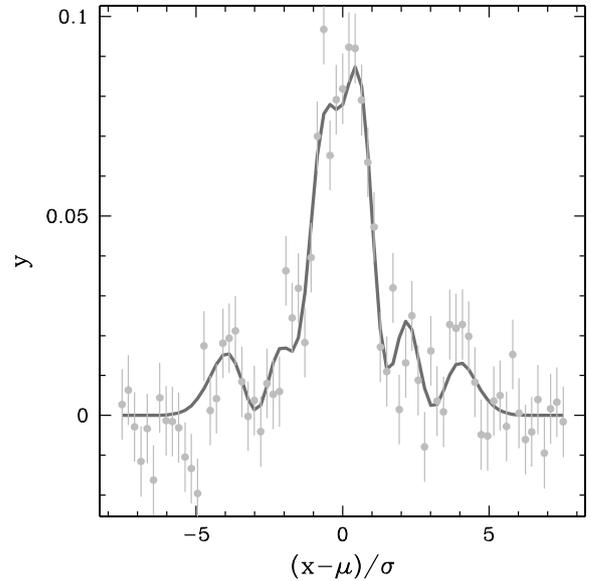


Figure 1. Example of simulated data $y_i = y(x_i)$ (equation 1). The solid grey line is the noise-free normalised generating process $y_0(x)$. The grey dots simulate a noisy measurement, i.e. they represent a data sample y obtained from y_0 by adding Gaussian noise (indicated by the error bars). The amount of noise is assumed to be constant along the x -axis. The SNR at the peak of the signal is 10. The generating input model is a Gauss–Hermite series up to order $n_{\text{GH}} = 10$ (cf. Table 1). We create $N_{\text{data}} = 71$ data points evenly spaced between $\pm 8\sigma$ of the Gauss component.

2 A TOY MODEL BASED ON HERMITE POLYNOMIALS

To illustrate the above outlined methods, we will try to recover a one-dimensional function $y(x)$ from noisy mock observations. To this end, we define $y_0(x)$ as a Gauss–Hermite series

$$y_0(x) = \frac{\gamma}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \times \left(1 + \sum_{i=3}^{n_{\text{GH}}} h_i H_i\left(\frac{x-\mu}{\sigma}\right)\right), \quad (1)$$

where H_i are Hermite polynomials.¹ The noisy mock data $y(x)$ are obtained by adding Gaussian noise. The highest non-zero order of our generating model is $n_{\text{GH}} = 10$ and the chosen coefficients are listed in Table 1. The resulting (noise-free) generating input model $y_0(x)$ is illustrated by the grey line in Fig. 1.

¹Any finite and suitably smooth function $F(x)$ with $\lim_{x \rightarrow \pm\infty} x^3 F(x) = 0$ for $x \rightarrow \pm\infty$ can be expanded into a Gauss–Hermite series (Myller-Lebedeff 1907). When (γ, μ, σ) equal the parameters of the best-fitting Gaussian function (we assume γ and σ are positive), then $h_1 = 1$ and $h_2 = 0$, and the series expansion can be written in the form of equation (1) (e.g. van der Marel & Franx 1993).

This toy model is inspired by the problem of fitting the shapes of gas emission lines in galaxies. In that case, y would represent a galaxy spectrum with the stellar continuum being subtracted and x would be the logarithm of the wavelength. Usually, the shape of such emissions is close to a Gaussian but one could interpret the input model as a very complex emission line with several more or less separated Gaussian-like sub-components. However, for the purposes of this work, one can also take this input model as an artificial mathematical model simply used to illustrate the capability of the model selection framework we propose.

3 TWO RECOVERY METHODS: PARAMETRIC VERSUS NON-PARAMETRIC FITS

We want to recover the generating model (Table 1) from the mock data in two different ways.

3.1 Parametric fits

As our first set of models $f(\Theta)$ we use the Gauss–Hermite series of equation (1) itself, i.e. we fit $f(\Theta) = y(x; \Theta)$. The parameters of this fit are $\Theta = (\gamma, \mu, \sigma, h_3, h_4, \dots, h_n)$. We test different Gauss–Hermite models by varying n_{GH} , the maximum order included in the series. In the following, only the case $n_{\text{GH}} \ll N_{\text{data}}$ will be relevant and we call these fits parametric fits. We will derive the best-fitting values of the $n_{\text{GH}} + 1$ free parameters $\Theta = (\gamma, \mu, \sigma, h_3, h_4, \dots, h_n)$ using a classical χ^2 minimization between the data y and the model $f(\Theta)$,

$$\chi^2 = \sum_{i=1}^{N_{\text{data}}} \frac{(y_i - f_i(\Theta))^2}{\epsilon_i^2}, \quad (2)$$

where $f_i(\Theta) = f(x_i; \Theta)$. Since the uneven Hermite polynomials are antisymmetric with respect to the y -axis and the even ones are symmetric, we will always increase n_{GH} in steps of two in our model fits.

3.2 Non-parametric fits

In addition to the parametric fits, we will also perform non-parametric fits where our model f consists of directly varying the $f_i \equiv \Theta_i$ of the signal at each of the $N_{\text{data}} = 71$ values of $(x_i - \mu)/\sigma$ where a data point has been simulated. Again, we will determine the 71 free parameters Θ_i of this model from a classical χ^2 minimization. The naive solution of this minimization problem would be trivial, as a non-parametric model can fit all the data (including the noise) perfectly when $f_i \equiv \Theta_i = y_i$, achieving $\chi^2 = 0$. Therefore, in this non-parametric case, we also include a penalty term to penalize our non-parametric model against arbitrarily unsmooth solutions. As penalty, we use the sum of the squared second derivatives of the normalised non-parametric model with respect to x , i.e. we minimize $\chi^2 + \alpha_S \mathcal{P}$ with

$$\mathcal{P} = \sum (f_{i+1} - 2f_i + f_{i-1})^2. \quad (3)$$

The factor α_S controls the strength of the smoothing constraints with $\alpha_S \rightarrow 0$ implying no smoothness enforced on the fits while $\alpha_S \rightarrow \infty$ implies strong smoothness constraints imposed on the model. By sampling different values of α_S , we construct different non-parametric models from which we can select. For both, the parametric as well as the non-parametric fits, we use a standard Levenberg–Marquardt routine for the parameter estimation.

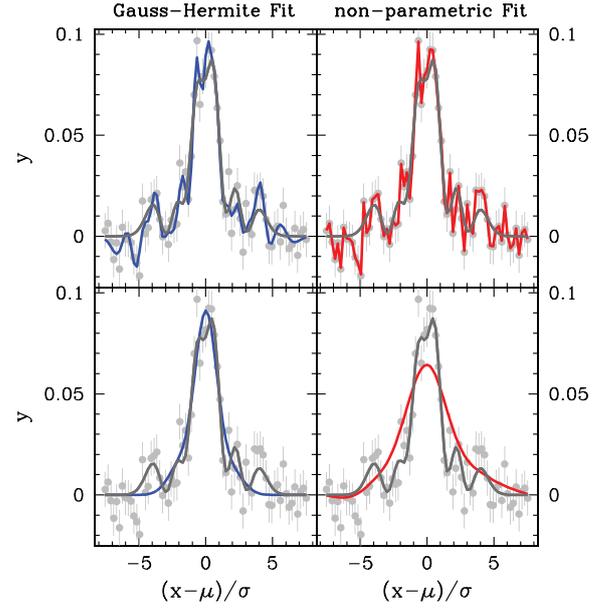


Figure 2. Example fits to the mock data of Fig. 1. The left-hand panels show parametric fits with too many polynomial components ($n_{\text{GH}} = 30$; top-left panel) and with too few polynomial components ($n_{\text{GH}} = 4$; bottom-left panel). The right-hand panels show non-parametric fits with a very low smoothing value ($\alpha_S = 10$; top-right panel) and a very high smoothing value ($\alpha_S = 3 \cdot 10^3$; bottom-right panel). The generating input model is shown by the grey line in each panel together with the noisy mock data realization (the same in each panel). The fits in the top panels do not recover the input model well because they *overfit* the data and follow the noise. The fits in the bottom panels do not recover the input model well either, but this time because they are *oversmooth* and not flexible enough to capture the structure of the generating model.

4 MODEL SELECTION – BALANCING OVERFITTING VERSUS UNDERFITTING

Fig. 2 shows four example fits to the mock data set presented in Fig. 1. The fits in the left-hand panels have been obtained with the parametric ansatz. In the top-left panel, $n_{\text{GH}} = 30$ – larger than in the generating model ($n_{\text{GH}} = 10$, cf. Section 2). As a consequence, the model overfits the data and is too structured, compared to the generating input model. The opposite is true for the fit shown in the bottom-left panel, where $n_{\text{GH}} = 4$. The resulting model is not flexible enough to capture all the structure of the generating model.

The panels on the right-hand side show the analogous cases for the non-parametric fits. In the top-right panel, the smoothing constraint is very weak such that the model fits the noise almost perfectly. In the bottom-right panel, the model is instead oversmooth and it cannot fit the data appropriately.

The goal of the following sections is to outline a simple method to find the optimum degree of model flexibility for both – the parametric and the non-parametric fit. The optimum degree of smoothness will be somewhere between the examples of overfitted and underfitted models shown in Fig. 2.

Model selection provides the general framework to choose between (not necessarily nested) models with different degrees of model flexibility. A commonly applied criterion used in this context is the AIC (Akaike 1974). If m is the number of free parameters in a fit and χ^2 the achieved goodness of fit, then a model is preferred when it has a lower

$$\text{AIC} = \chi^2 + 2m. \quad (4)$$

The AIC originally comes from information theory and provides optimal model recovery in the sense that it aims to minimize the (estimated) Kullback–Leibler divergence – a measure of the expected information loss – between model and the actual generating process (cf. Section 6). Intuitively, the AIC selects the model with the smallest number of variables m among all models that provide a statistically viable fit, i.e. among all models with $\chi^2 + m \approx N_{\text{data}}$. This is because for all these viable models, the AIC roughly reduces to $\text{AIC} = N_{\text{data}} + m$ and, hence, becomes smallest for the lowest m . As such it's akin to Occam's razor.

Our intuitive understanding that the information loss in a fit tends to increase with the number of fitted parameters is often – but not always – sufficient to select an optimal model. An obvious requirement is that the number of parameters m has to be known. This seems to be trivial for parametric fits where one usually knows every fit parameter explicitly and m can be determined by simply counting the variables of the fitting function. However, even in the parametric case the determination of m can become non-trivial when the model is strongly non-linear or when the parameters are not allowed to vary arbitrarily but instead are subject to constraint equations (Andrae, Schulze-Hartung & Melchior 2010).

The same applies to non-parametric models. Here, even more complications arise from smoothing penalties that often depend on one or several continuous smoothing parameters. Naively, the stronger the imposed smoothing constraints, the less flexible the model becomes. The invariable number m of fitted parameters can therefore no longer be a measure of a model's flexibility or responsiveness to noise, respectively. This is important because noise – by definition – does not carry any information about the data generating structure. Hence, it is actually a model's responsiveness to noise that determines the expected information loss. And if this is no longer encoded in m then the simple model selection criterion of equation (4) will no longer be sufficient. In fact, when penalties with a continuous parameter are present, the *effective* number of variables is expected to become a continuous variable as well. For example, in the above smoothing case: the stronger the smoothing the less flexible the model and the smaller the effective number of variables. In the next two sections, we will demonstrate that the selection criterion equation (4) can still be used under such circumstances, provided, however, that the generalized concept of the *effective* number of parameters is applied to quantify a model's flexibility.

5 QUANTIFYING MODEL FLEXIBILITY – THE NUMBER OF EFFECTIVE FREE PARAMETERS

In Lipka & Thomas (2021), we first introduced such a generalized concept of *effective* number of free parameters by using a flexible bootstrap method to estimate them. The following section is a brief review of the concepts first shown in Lipka & Thomas (2021).

Strictly speaking the number of free parameters m is well defined only for parametric models that depend linearly on their free parameters (e.g. Hastie, Tibshirani & Friedman 2013) and have no a priori constraints imposed on their parameters (e.g. Andrae et al. 2010). However, the concept of free parameters can be generalized formally to more complex statistical models without relying on such restricting assumptions about the underlying model structure (e.g. Ye 1998). In such generalized frameworks, the resulting degrees of freedom (i.e. the actual model flexibility) typically differs significantly from the number of variables of the fit model, and thus one cannot derive the model flexibility by simply counting the number of variables.

Therefore, we employ bootstrap iterations to estimate the *effective* number of free parameters (in the following m_{eff}). To this end, we

establish a number of N_{boot} bootstrap data sets z for each fit model by adding random Gaussian noise (based on the observed noise estimate ϵ) to an initial fit $f(\hat{\Theta}_y)$ of said model to the observed data sample y . Thus, a set of bootstrap data z is generated at every data point i by $z_i = f(\hat{\Theta}_y) + \mathcal{N}(0, \epsilon_i)$, where $\mathcal{N}(0, \epsilon_i)$ is a Monte Carlo realization drawn from the Gaussian distribution with mean 0 and standard deviation ϵ_i . The goal of this bootstrap resampling technique is to emulate (or redraw) the observed data sample. As such bootstrapping assumes that the initial fit $f(\hat{\Theta}_y)$ represents the (noise-free) data generating process well enough such that the bootstrap data can be treated as a resample of the observed sample.² Each of the N_{boot} sets of bootstrap data z^κ (for $\kappa = 1, \dots, N_{\text{boot}}$) is then fitted by the same model for which one attempts to estimate the flexibility, denoted as $f_i(\hat{\Theta}_z^\kappa)$.

The flexibility m_{eff} should be a measure of responsiveness of the model fit to noisy data. As such, a more flexible model should be able to follow more of the deviations in the data that were induced by the bootstrap noise than a less flexible model can. Therefore, a 'natural' measure of this responsiveness is the normalized correlation of the model fit and the noisy data it was fitted to³

$$m_{\text{eff}} = \sum_{i=1}^{N_{\text{data}}} \frac{\text{Cov}(f_i(\hat{\Theta}_z), z_i)}{\epsilon_i^2} \leq N_{\text{data}}. \quad (5)$$

Motivated by this expression (5), we introduce

$$m_{\text{eff}}^\kappa = \sum_{i=1}^{N_{\text{data}}} \left(\frac{f_i(\hat{\Theta}_z^\kappa) - f_i(\hat{\Theta}_y)}{\epsilon_i} \right) \left(\frac{z_i^\kappa - f_i(\hat{\Theta}_y)}{\epsilon_i} \right), \quad (6)$$

where the sum goes over all data points N_{data} and κ is the index of the bootstrap iteration. To reduce dependence on the specific noise realizations of the bootstrap data, one should ideally average the results of equation (6) over multiple bootstrap iterations $\kappa = 1, \dots, N_{\text{boot}}$ such that

$$m_{\text{eff}} = E(m_{\text{eff}}^\kappa) \simeq \frac{1}{N_{\text{boot}}} \sum_{\kappa=1}^{N_{\text{boot}}} m_{\text{eff}}^\kappa. \quad (7)$$

Of course multiple iterations can be very computationally expensive if the fitting procedure in itself is complex. Fortunately, one can already achieve decent results without the need of that many bootstrap iterations as we will discuss in Section 8.

For linear models that are not subject to a penalty and for which the parameters are not restricted by equality/inequality constraints, the number m of fitted parameters happens to equal m_{eff} and therefore a simple count of the number of parameter m is a viable measure for the responsiveness of the model to noise. We show this explicitly in Appendix A. Note that in standard linear theory m actually only equals the naive number count of parameters if all parameters are linearly independent (cf. Hastie et al. 2013). Nevertheless, the equality of m_{eff} and m (given aforementioned preconditions) still holds for dependent parameters as long as m is calculated using standard linear theory instead of a naive counting of the number of parameters.

²If this approximation is too crude, e.g. if the initial fit model is very different from the actual data generating process, then the bootstrapping estimation of m_{eff} is not legitimate. However, such models can easily be rejected anyways due to their overall bad fit to the data (e.g. if $\chi^2/N_{\text{data}} \gg 1$).

³Equation (5) can be shown to be equivalent to the definition generalized degrees of freedom of Ye (1998), which is a formal extension of the number of free parameters to non-linear models.

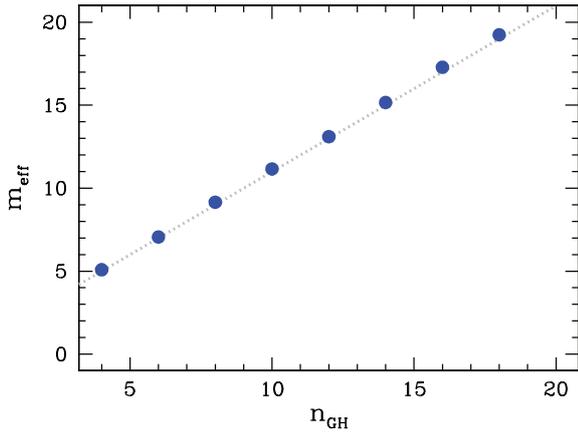


Figure 3. Estimated number of effective parameters m_{eff} for the parametric Gauss–Hermite models described in Section 3.1. The blue dots show the numerically derived m_{eff} with $N_{\text{boot}} = 500$. For the Gauss–Hermite models, one can simply count the number of free parameters to $m_{\text{eff}} = n_{\text{GH}} + 1$ (dotted line). All the blue points fall exactly on the dotted line: the numerical evaluation of m_{eff} is very accurate.

Our non-parametric fits provide an example for such a linear model (with $\partial f_i / \partial \Theta_j = \delta_{ij}$) if no smoothing penalty is applied ($\alpha_S = 0$). In that case, the model will always yield $\chi^2 = 0$ with $f_j \equiv \Theta_j = y_j$ such that

$$m_{\text{eff}} = \sum_{i=1}^{N_{\text{data}}} \frac{\text{Cov}(z_i, z_i)}{\epsilon_i^2} = N_{\text{data}} = m, \quad (8)$$

as expected.

For non-linear models, the equality $m_{\text{eff}} = m$ might not hold in general, but under appropriate regularity conditions it will certainly hold locally. In practice, models that change their properties significantly over the uncertainty region of the data will in most cases not be very useful. Hence, if the model changes only slowly over the region in data space sampled by the bootstraps then $m_{\text{eff}} = m$ will still hold (in the absence of penalties or parameter constraints). An example is the Gauss–Hermite series of Section 3.1 where the parameter vector $\Theta = (\gamma, \mu, \sigma, h_3, h_4, \dots, h_n)$ has a well-defined number of $n_{\text{GH}} + 1$ elements, or free parameters respectively. Fig. 3 shows the numerically obtained m_{eff} of these non-linear models versus the (counted) number of free parameters. As expected, after $N_{\text{boot}} = 500$ bootstraps the estimated effective number of parameters is equivalent to the counted number of free parameters.

In model selection, the goal is to rank different models according to the estimated relative information loss between them. The classical $\text{AIC} = \chi^2 + 2m$ has been shown to be an unbiased estimator of this information loss, e.g. in the context of maximum-likelihood fits. However, we have motivated above that for penalized models, e.g. the ability of the model to adapt to noise depends on the strength of the penalty term and cannot be expressed by the invariable number m . By construction, m_{eff} is a more general measure of the model’s flexibility, independent of the presence of a penalty or other constraints on the parameters. Therefore, we argue that model selection in a more general context should involve m_{eff} rather than m (the latter only being a measure of the responsiveness to noise under certain circumstances). Section 6 is dedicated to such a generalized model selection that extends to penalized models and, as such, will involve m_{eff} rather than m .

6 PENALIZED LIKELIHOOD: MODEL SELECTION WITH EFFECTIVE NUMBER OF PARAMETERS

In Lipka & Thomas (2021), within the context of orbit superposition models for galaxies, we have tested model selection techniques using m_{eff} rather than m . For a fixed gravitational potential, the said orbit models are linear, but have a non-linear (maximum-entropy) penalty in our implementation (Richstone & Tremaine 1988; Thomas et al. 2004). Comparing different weighting schemes $\chi^2 + w_m m_{\text{eff}}$, we found that the AIC analogue (i.e. $w_m = 2$) performs best in estimating underlying properties of the data-generating processes. Therefore, the results of Lipka & Thomas (2021) suggest that the AIC can be generalized to penalized models by the substitution $m \rightarrow m_{\text{eff}}$.

In the following, we will motivate how model selection can be generalized for penalized models in a more formal way. For the reader who wants to skip this rather technical discussion, we preempt the important results of this section: for penalized models, model selection indeed consists of minimizing the generalized $\text{AIC}_p = \chi^2 + 2m_{\text{eff}}$. We derive this exactly for linear models. In Section 7, we will apply the generalized AIC_p to the toy model introduced in Section 2.

It is out of the scope of this paper to give a complete introduction to the foundation of model selection. A very good overview can be found in Burnham & Anderson (2002). We simply start by recalling that Akaike model selection consists of minimizing the *expected, estimated* information loss

$$-E_y E_z (\log \mathcal{L}(z | \hat{\Theta}_y)) \quad (9)$$

(e.g. Chapter 7.2, Burnham & Anderson 2002). Here, $\log \mathcal{L}(z | \hat{\Theta}_y)$ is the logarithm of the likelihood \mathcal{L} of some fictitious data z (see below) at the maximum-penalized likelihood estimate $\hat{\Theta}_y$ of the model parameters Θ . The double expectation E_y and E_z deserve some further comments. The outer expectation E_y is meant to reflect that – conceptionally – we aim at minimizing the *expected* information loss over large samples of actual data y . In the context of our toy model (Section 2), this would correspond to averaging the results over several mock data sets (we will come back to this in Section 8.4). However, in most practical applications one has only a single data set and needs an unbiased estimate of the information loss based on the actual data at hand (see below). The inner expectation E_z reflects that the Kullback–Leibler divergence – the measure of the information loss that underlies equation (9) – is an integral that happens to have the *form* of an expectation value. Hence, the integral over the integration variable z can be expressed and interpreted as an expectation value over some fictitious data samples z . Below it will turn out that the bootstrap iterations we introduced in Section 5 are effectively the computation of the inner expectation E_z over this fictitious data sample z of equation (9).

In the following – for simplicity – we restrict ourselves to a situation where the ‘truth’ corresponds to one model among our candidates, i.e. there is a parameter vector Θ_0 of the ‘true’ values of Θ .⁴ If we would know Θ_0 we could use it in equation (9) to calculate the true information loss. However, in reality we only have an estimate $\hat{\Theta}_y$ based on some noisy data y . Therefore, equation (9) only quantifies the *expected, estimated* information loss based on $\hat{\Theta}_y$.

⁴Model selection does not depend on the true model being among the candidates. While this complicates the discussion (e.g. Burnham & Anderson 2002), it does not change the conclusions in our context.

The standard derivation of AIC starts with the Taylor expansion

$$\log \mathcal{L}(z|\hat{\Theta}_y) \approx \log \mathcal{L}(z|\Theta_0) + \left[\frac{\partial \log \mathcal{L}(z|\Theta_0)}{\partial \Theta} \right]^T (\hat{\Theta}_y - \Theta_0) + \frac{1}{2} (\hat{\Theta}_y - \Theta_0)^T \frac{\partial^2 \log \mathcal{L}(z|\Theta_0)}{\partial \Theta^2} (\hat{\Theta}_y - \Theta_0), \quad (10)$$

around this true parameter vector Θ_0 . Here and in the rest of the paper, we use the compact notation $\partial^2 \log \mathcal{L}(z|\Theta_0)/\partial \Theta^2$ to denote the Hessian matrix of $\log \mathcal{L}$ evaluated at Θ_0 . For maximum-likelihood models, the expectation of the linear term obviously vanishes. In our case of linear penalized models, this holds true since $E_z(z) = f(\Theta_0)$ (cf. Appendix B). Even if the truth is not among the candidate models this holds true under very weak conditions because $E_y(\hat{\Theta}_y) = \Theta_0$.

Then, using a second Taylor expansion the unknown $\log \mathcal{L}(z|\Theta_0)$ in equation (10) is approximated as

$$\log \mathcal{L}(z|\Theta_0) \approx \log \mathcal{L}(z|\hat{\Theta}_z) + \left[\frac{\partial \log \mathcal{L}(z|\hat{\Theta}_z)}{\partial \Theta} \right]^T (\Theta_0 - \hat{\Theta}_z) + \frac{1}{2} (\Theta_0 - \hat{\Theta}_z)^T \frac{\partial^2 \log \mathcal{L}(z|\hat{\Theta}_z)}{\partial \Theta^2} (\Theta_0 - \hat{\Theta}_z). \quad (11)$$

Within the classical maximum-likelihood framework the linear term vanishes exactly for each z – by construction. The expectations of the remaining second-order terms of equations (10) and (11) can then be shown to combine to (the negative of) the number of fitted parameters m which then leads to the classical form of the AIC (e.g. Burnham & Anderson 2002).

In contrast, within a *penalized* maximum-likelihood framework, the linear term does not vanish and needs to be taken into account when combining equations (10) and (11). For linear models, a simplification arises from the fact that $\partial^2 \log \mathcal{L}/\partial \Theta^2$ is a constant. Furthermore, since we will take the double expectation $E_y E_z$ and $E_y E_z(h(y)) = E_y E_z(h(z))$ for any function h , we can substitute $\hat{\Theta}_y$ by $\hat{\Theta}_z$ in equation (10). Combining all the above, equation (9) becomes

$$E_y E_z(\log \mathcal{L}(z|\hat{\Theta}_y)) \approx E_y E_z(\log \mathcal{L}(z|\hat{\Theta}_z)) + E_y E_z(h(z|\hat{\Theta}_z)), \quad (12)$$

with

$$h(z|\hat{\Theta}_z) = \left[\frac{\partial \log \mathcal{L}(z|\hat{\Theta}_z)}{\partial \Theta} \right]^T (\Theta_0 - \hat{\Theta}_z) + (\Theta_0 - \hat{\Theta}_z)^T \frac{\partial^2 \log \mathcal{L}(z|\hat{\Theta}_z)}{\partial \Theta^2} (\Theta_0 - \hat{\Theta}_z). \quad (13)$$

In our bootstrap simulations, we use the estimate $\Theta_0 \approx \hat{\Theta}_y$. Then, using equations (B19) and (B20), we find

$$E_z(h(z|\hat{\Theta}_z)) = E_z \left[(z - f(\hat{\Theta}_z))^T \Sigma^{-1} (f(\hat{\Theta}_y) - f(\hat{\Theta}_z)) + (f(\hat{\Theta}_y) - f(\hat{\Theta}_z))^T \Sigma^{-1} (f(\hat{\Theta}_y) - f(\hat{\Theta}_z)) \right], \quad (14)$$

and it is straightforward to show that this expectation value equals (the negative of) m_{eff} (cf. equation B21).

In the common situation where one cannot perform the expectation $E_y(\log \mathcal{L}(y|\hat{\Theta}_y))$ or, equivalently, $E_z(\log \mathcal{L}(z|\hat{\Theta}_z))$ one uses the estimate $\log \mathcal{L}(y|\hat{\Theta}_y)$, which for Gaussian errors reads $-\chi^2/2$. Taking everything together, model selection under penalized likelihood conditions then consists of minimizing $\chi^2/2 + m_{\text{eff}}$ or, equivalently

$$\text{AIC}_p = \chi^2 + 2m_{\text{eff}}. \quad (15)$$

It is natural to assume that the extended criterion of equation (15) also holds (at least locally) for more general, non-linear models

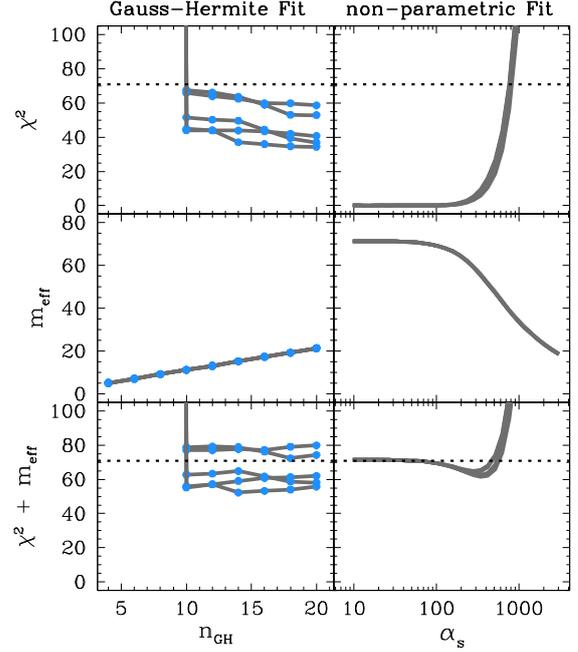


Figure 4. Illustration of varying model complexity in the Gauss–Hermite (parametric) fits (left-hand panels) and the non-parametric fits (right-hand panels). The generating model is the Gauss–Hermite series with $n_{\text{GH}} = 10$ shown as the solid curve in Fig. 1. Each panel shows five grey lines, one for each of five different mock realizations of noisy data similar to the data shown in Fig. 1 but assuming an SNR of 100 at the peak of the model. From top to bottom the plot shows the goodness-of-fit χ^2 , the number of effective parameters m_{eff} and $\chi^2 + m_{\text{eff}}$. The number of data points, $N_{\text{data}} = 71$, is illustrated by the horizontal dotted lines. A statistically viable model must have a $\chi^2 + m_{\text{eff}} \approx N_{\text{data}}$. n_{GH} is a discrete parameter of the parametric fits, its actually allowed values are highlighted in blue, the grey lines connecting the points have been added to better illustrate the trend. In the non-parametric case, α_S is a differentiable parameter and m_{eff} is a continuous function. For the parametric fits, the model flexibility increases with the maximum order of the Gauss–Hermite fit, n_{GH} . For $n_{\text{GH}} \geq 10$, all fits become statistically viable. In the non-parametric case, the flexibility decreases as a function of the smoothing parameter α_S , together with m_{eff} . Below $\alpha_S \lesssim 3 \cdot 10^2$, all models lead to statistically acceptable fits. The middle-left panel confirms $m_{\text{eff}} = n_{\text{GH}} + 1$, i.e. the expected number of parameters for the parametric case.

under appropriate regularity conditions (which to derive is out of the scope of this paper). In fact, in the next Section 7, we will see that AIC_p works equally well for our linear non-parametric model and non-linear Gauss–Hermite model. In the absence of a penalty (or of constraint equations for the parameters), $m_{\text{eff}} = m$ (Section 5) and $\text{AIC}_p = \text{AIC}$.

We note that the difference between AIC and AIC_p – i.e. the replacement $m \rightarrow m_{\text{eff}}$ – arises because we do not assume that the models obey the maximum-likelihood condition in our derivation of AIC_p as is the case for the classical AIC. Therefore, AIC_p encompasses penalized models as well. However, we want to stress that our derivation is independent of the specific conditions for *penalized* maximum-likelihood models, meaning AIC_p is not restricted to these models and may be applied in an even more general sense.

7 EXAMPLE RESULTS

Fig. 4 shows the result for both the parametric method (left-hand panels) and the non-parametric method (right-hand panels) in terms of fits to five different mock data realizations. In the parametric case, as

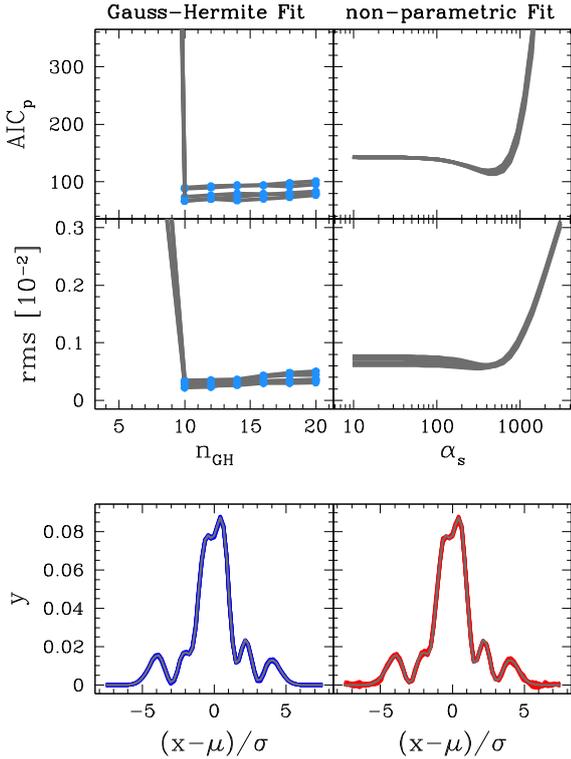


Figure 5. The same models as in Fig. 4. The top panels show the $AIC_p = \chi^2 + 2m_{\text{eff}}$. The middle panels show the rms difference between the generating input model and the best-fitting reconstruction of this input model from the fit to the mock data. Finally, the bottom panels show the generating model (grey) and the fitted reconstructions (blue/red) at the minimum AIC_p . The reconstruction of the input model is extremely good in both cases. In the parametric and non-parametric cases, the AIC_p selection yields the model with the smallest rms difference to the generating input model.

expected, the number of fit variables ($n_{GH} + 1$) increases with n_{GH} and therefore the χ^2 decreases with n_{GH} (top-left and middle-left panels). In fact, as long as $n_{GH} < 10$ – i.e. when the number of fitted variables is smaller than in the generating model – the fits do not yield a statistically viable fit to the data, because $\chi^2 + m_{\text{eff}} \gg N_{\text{data}}$ (bottom-left panel). For $n_{GH} \geq 10$, while χ^2 continues to decrease with increasing n_{GH} , all fits actually provide statistically equivalent representations of the data as $\chi^2 + m_{\text{eff}}$ stays roughly constant in this regime. Implying the improvement in the goodness-of-fit χ^2 is not significant but just as large as expected from the increased model flexibility.

The non-parametric fits (right-hand panels) behave similar, though m_{eff} increases from left to right (opposite to the parametric case) in this representation of α_s . For low values of α_s , the models are essentially unaffected by the smoothing penalty. As a result, each model variable θ_i becomes an entirely independent model variable and $m_{\text{eff}} = N_{\text{data}}$. In this regime, the model adapts perfectly to the noise in the data resulting in $\chi^2 \rightarrow 0$. The larger α_s the stronger the smoothing constraints become, meaning that m_{eff} continuously decreases whereas χ^2 accordingly increases. Over a large interval of α_s this happens at constant $\chi^2 + m_{\text{eff}} \approx N_{\text{data}}$. That is, while the model becomes smoother it still leads to a statistically viable representation of the data. At some point, however, the smoothing constraints become so dominant that χ^2 and $\chi^2 + m_{\text{eff}}$ increase significantly above N_{data} . In that case, the model becomes so dominated by the smoothing function that it cannot yield a good representation of the data anymore. A noticeable difference between the parametric and

the non-parametric fits is that the non-parametric χ^2/m_{eff} curves appear to be much smoother than the parametric fits. We will come back to this in Section 8.4.

Fig. 5 shows the AIC_p (top panels), the recovery of the input model at the lowest AIC_p (bottom panel), and the rms (root-mean-square) between the generating input model and recovery from the fit (middle panels) – again for both the parametric and the non-parametric fits. For the rms, we sum over the squared differences between the generating model and the fit at the N_{data} argument values of the data points. Unsurprisingly, the AIC_p of the parametric fits has a minimum at $n_{GH} = 10$, the value used for the input model. For larger n_{GH} , even though the goodness-of-fit χ^2 improves, the AIC_p increases again because the models do not lead to a *significantly* better fit. This behaviour of the AIC_p is mirrored exactly by the rms. The fact that the rms worsens with increasing n_{GH} even though fits with $n_{GH} > 10$ are statistically viable is due to the fact that the models adapt more and more to the noise in the data, i.e. they start to overfit. The recovered model at the optimum n_{GH} agrees very well with the input model.

The good recovery of the input model with the parametric fits is not that surprising since the data generating model (equation 1) is among the candidate models in this case. However, the recovery with the non-parametric models is almost equally good (right-hand panels) even though in this case the generating model is not among the candidates. Again, the change of rms and the behaviour of the AIC_p are very similar and the minimum AIC_p is found to be where the recovery of the model is best. This shows that one can identify the optimal smoothing, or equivalently the optimum number of fit variables, even in the non-parametric case using the AIC_p optimization and our definition of *effective* free parameters (Section 5).

These above results and the fact that one can identify the model with the optimum number of fit variables from the AIC_p and the m_{eff} does not depend on the assumed signal-to-noise ratio (SNR). For the above fits, the SNR at the peak of the model was set to $\text{SNR} = 100$. Fig. 6 shows the results for $\text{SNR} = 10$. While the recovery of the model gets more difficult due to the increased noise in the data, one can still identify the model with the optimum degree of model flexibility with the above-described method.

8 EVALUATING THE EFFICIENCY

It is common in non-parametric models to calibrate the smoothing by use of Monte Carlo simulations. With the above bootstrap approach such simulations are not necessary anymore. However, the efficiency of this approach will depend on the number of bootstrap iterations N_{boot} necessary to obtain an accurate estimate of the optimum α_s .

In the AIC_p framework, the best choice for α_s follows from

$$\frac{dAIC_p}{d\alpha_s} = 0 \quad (16)$$

or

$$\frac{d\chi^2}{d\alpha_s} = -2 \frac{dm_{\text{eff}}}{d\alpha_s}, \quad (17)$$

respectively. Therefore, we actually only need an accurate estimate for $dm_{\text{eff}}/d\alpha_s$, rather than for m_{eff} itself, to determine the optimum degree of smoothing α_s .

In this section, we will compare the scatter in m_{eff} itself and the scatter of its derivative $dm_{\text{eff}}/d\alpha_s$ with the goal to predict the number of bootstraps iterations N_{boot} required to find the optimum α_s .

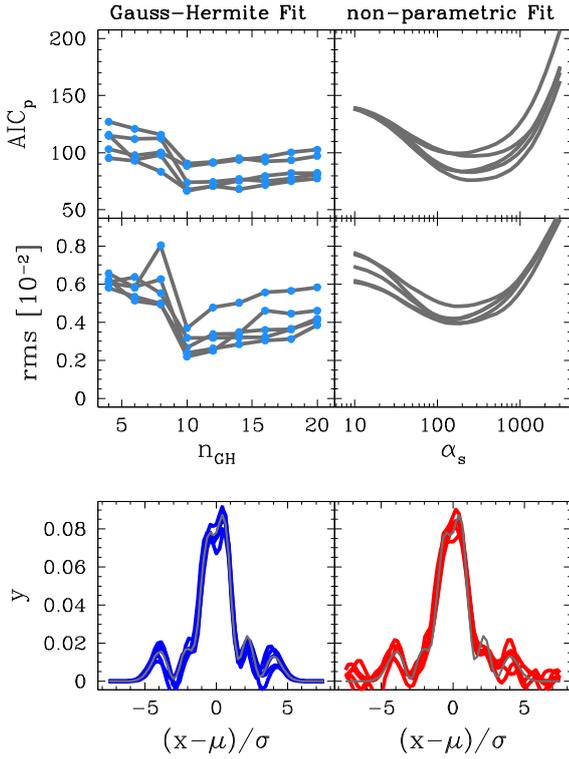


Figure 6. Same as Fig. 5 but the data are significantly noisier with an SNR = 10.

8.1 Bootstrap scatter in m_{eff}

To estimate the scatter in m_{eff} it is convenient to define

$$a_i = \frac{f_i(\hat{\Theta}_z^k) - f_i(\hat{\Theta}_y)}{\epsilon_i} \quad (18)$$

and

$$b_i = \frac{z_i^k - f_i(\hat{\Theta}_y)}{\epsilon_i} \quad (19)$$

and treat them as random variables. The products $c_i = a_i b_i$ define the individual contributions to m_{eff}^k (equation 6) such that

$$\text{Var}(m_{\text{eff}}^k) = \text{Var}\left(\sum_i^{N_{\text{data}}} c_i\right). \quad (20)$$

From stochastic theory, we can use Bienaymé's identity:

$$\text{Var}\left(\sum_i^{N_{\text{data}}} c_i\right) = \sum_i^{N_{\text{data}}} \text{Var}(c_i) + \sum_{i \neq j}^{N_{\text{data}}} \text{Cov}(c_i, c_j) \quad (21)$$

and

$$\text{Var}(c_i) = \text{Cov}(a_i^2, b_i^2) - [\text{Cov}(a_i, b_i) + E(a_i)E(b_i)]^2 + (\text{Var}(a_i) + E(a_i)^2)(\text{Var}(b_i) + E(b_i)^2) \quad (22)$$

to evaluate equation (20). By construction, $E(b_i) = 0$ and $\text{Var}(b_i) = 1$ such that equation (22) simplifies to

$$\text{Var}(c_i) = \text{Cov}(a_i^2, b_i^2) - \text{Cov}(a_i, b_i)^2 + \text{Var}(a_i^2). \quad (23)$$

The quantities a_i and b_i (and c_i , respectively) can be calculated during the bootstrap iterations. After all N_{boot} bootstrap iterations, the scatter

of m_{eff} is

$$\text{Var}(m_{\text{eff}}) = \text{Var}(m_{\text{eff}}^k) / N_{\text{boot}} \quad (24)$$

and can be estimated using the above equations.

The left-hand panels of Fig. 7 illustrate estimates of m_{eff} based on different N_{boot} (solid lines) together with the calculated scatter $\Delta m_{\text{eff}} = \sqrt{\text{Var}(m_{\text{eff}})}$ (dotted lines). We also include m_{eff} for $N_{\text{boot}} = 2500$ as a reference (dashed line). Both, the estimate for m_{eff} and for its scatter improve with increasing N_{boot} such that they can be determined with any desired accuracy. In practice, already after a relatively low number of bootstrap iterations $N_{\text{boot}} \sim 10$ the scatter from equation (24) captures the uncertainty in m_{eff} very well and can be used to estimate the required N_{boot} .

The exact behaviour of m_{eff} and Δm_{eff} will depend on the model function, the data, and the smoothing function. However, in many cases – as in Fig. 7 – the scatter Δm_{eff} can be presumed to increase with α_s . Especially when the smoothing function biases the fit towards a single unique solution. If this favoured reference model is not well chosen and far away from the true generating model (compared to ϵ_i), the Δm_{eff} will be dominated by the $E(a_i^2)$ term (cf. equation 23) leading to a large scatter in m_{eff} .

In our case, the smoothing function does not prefer a single unique set of values for the fitted f_i because any straight line with any combination of slope/intercept will minimize the penalty function. Still, the scatter in m_{eff} increases noticeably with α_s . As a rule of thumb we found that the scatter typically grows with $\chi^2(\alpha_s)$, i.e. the goodness of fit of the original model $f(\hat{\Theta}_y)$ at α_s . Fortunately, this also means that the regions with the largest scatter are typically not of interest anyways.

8.2 Bootstrap scatter in $dm_{\text{eff}}/d\alpha_s$

As stated above, for the AIC_p optimization of α_s the scatter in $dm_{\text{eff}}/d\alpha_s$, and not in m_{eff} , is the more important quantity. Therefore, we will now evaluate the scatter in the derivative of m_{eff} with respect to α_s .

Suppose we have two estimates of m_{eff} at two neighbouring values of α_s , $m_{\text{eff}}(\alpha_s)$ and $m_{\text{eff}}(\alpha_s + d\alpha_s)$. The variance of the difference $dm_{\text{eff}}^k = m_{\text{eff}}^k(\alpha_s + d\alpha_s) - m_{\text{eff}}^k(\alpha_s)$ is

$$\text{Var}(dm_{\text{eff}}^k) = \text{Var}(m_{\text{eff}}^k(\alpha_s)) + \text{Var}(m_{\text{eff}}^k(\alpha_s + d\alpha_s)) - 2 \text{Cov}(m_{\text{eff}}^k(\alpha_s), m_{\text{eff}}^k(\alpha_s + d\alpha_s)) \quad (25)$$

and, in analogy to m_{eff} ,

$$\text{Var}(dm_{\text{eff}}) = \text{Var}(dm_{\text{eff}}^k) / N_{\text{boot}}. \quad (26)$$

What matters here – beyond the scatter of m_{eff} itself – is the correlation or covariance between neighbouring fits. For our penalty function (equation 3), the curves of the individual $m_{\text{eff}}^k(\alpha_s)$ are a smooth function of α_s (cf. Fig. 7). This is ensured if the penalty function is differentiable in α_s , because then the correlation between neighbouring models is nearly maximal. For example, if a specific noise pattern led to a bootstrap data set z that happened to result in a relative large m_{eff}^k at α_s [compared to the mean $m_{\text{eff}}(\alpha_s)$], then this will very likely also be true for $m_{\text{eff}}^k(\alpha_s + d\alpha_s)$. This holds locally if one uses the *same* noise pattern for neighbouring models. When comparing bootstrap fits at sufficiently different α_s this correlation will be weaker or might disappear completely (i.e. some of the m_{eff}^k of Fig. 7 cross).

In the middle column of panels of Fig. 7, we plot $dm_{\text{eff}}(\alpha_s)$ for different N_{boot} together with $\Delta dm_{\text{eff}} = \sqrt{\text{Var}(dm_{\text{eff}})}$. Due to the high degree of correlation between fits (and m_{eff}) at neighbouring α_s , the

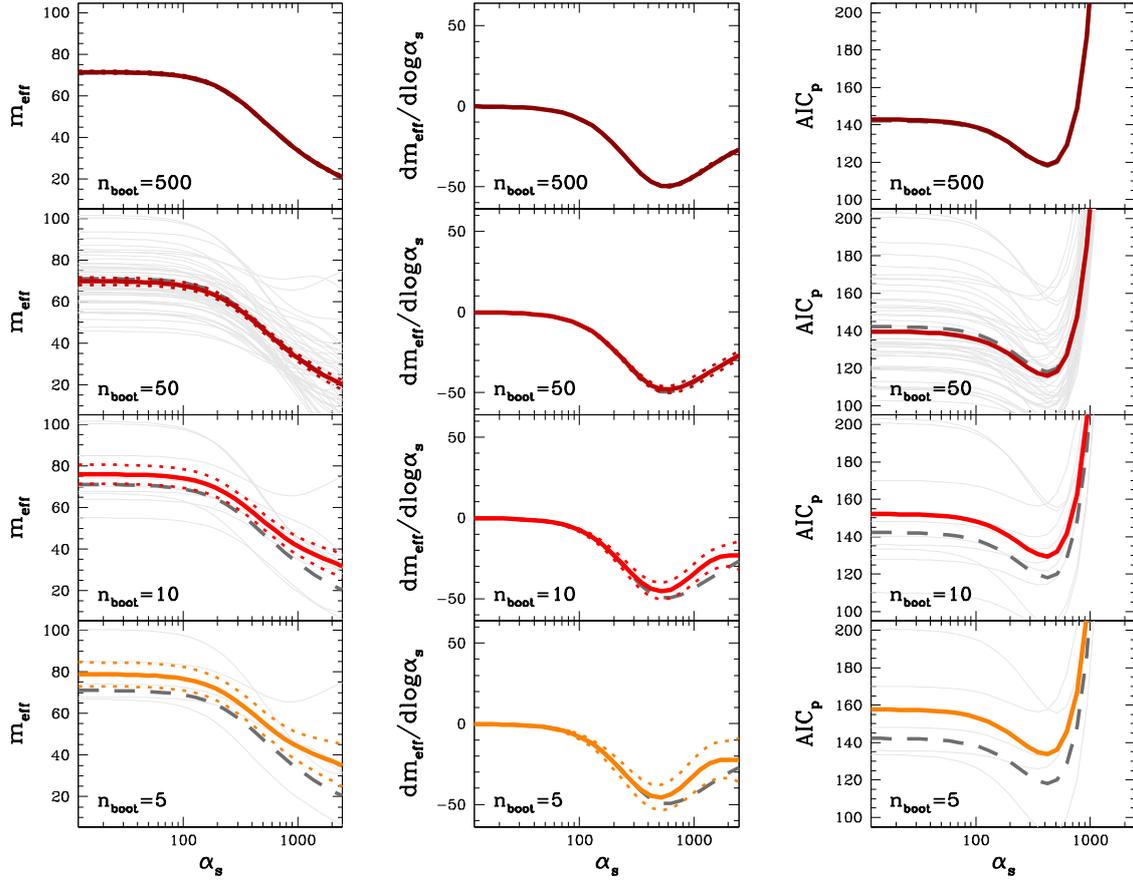


Figure 7. Scatter analysis of the models presented in Figs 4 and 5. Left-hand panels: Number of effective parameters m_{eff} as a function of the smoothing factor α_S for different N_{boot} (labelled in each panel). The coloured solid curves show the mean m_{eff} over the respective N_{boot} and the dotted lines indicate Δm_{eff} . For reference, the grey dashed line represents the case for $N_{\text{boot}} = 2500$. In the three panels with $N_{\text{boot}} \leq 50$, the individual m_{eff}^k from each individual bootstrap iteration are included as well (thin solid lines). Most of the scatter in m_{eff} comes from individual m_{eff}^k being scattered in the vertical direction as a whole in response to the particular noise pattern of each bootstrap iteration. The behaviour $m_{\text{eff}}^k(\alpha_S)$ depends very little on the noise pattern, in particular locally. Middle panels: Similar as the left-hand panels but the derivative $dm_{\text{eff}}/d\log\alpha_S$ is shown instead of m_{eff} . Rather than m_{eff} it is this derivative that is crucial to find the minimum of AIC_p . Because the scatter in m_{eff} mostly results from vertical shifts in the entire curves $m_{\text{eff}}^k(\alpha_S)$, the derivative $dm_{\text{eff}}/d\log\alpha_S$ is very easy to compute with a higher accuracy, even with small N_{boot} . Right-hand panels: AIC_p as a function of α_S for different N_{boot} . The vertical dotted line indicates the value of α_S where the rms between the generating input model and the reconstruction fit is smallest. For all the shown N_{boot} it is correctly recovered by the minimum of AIC_p . In many situations even a single bootstrap iteration can be enough to get a decent optimization of the smoothing in a non-parametric fit.

scatter in dm_{eff} is very much reduced and even with less than $N_{\text{boot}} < 10$ bootstraps one can identify the characteristic behaviour of dm_{eff} .

Finally, the panels at the very right of Fig. 7 show $\text{AIC}_p(\alpha_S)$ for different N_{boot} . For the optimization of α_S , the $\chi^2(\alpha_S)$ term is significant as well. Similarly to m_{eff} , we have

$$\begin{aligned} \text{Var}(d\chi_k^2) &= \text{Var}(\chi_k^2(\alpha_S)) + \text{Var}(\chi_k^2(\alpha_S + d\alpha_S)) \\ &\quad - 2 \text{Cov}(\chi_k^2(\alpha_S), \chi_k^2(\alpha_S + d\alpha_S)) \end{aligned} \quad (27)$$

and basically all the above considerations about m_{eff} can be taken over to $\chi^2(\alpha_S)$. As long as the penalty is differentiable in α_S the function $\chi^2(\alpha_S)$ will be smooth (its actual behaviour is shown in the top-right panel of Fig. 4). Consequently, it is not surprising that the AIC_p curves are very smooth even for very small N_{boot} .

The vertical dotted line in the right-hand panels of Fig. 7 indicates the value of the smoothing factor α_S where the rms between the generating model and the fit has its minimum, i.e. the best model. In all the cases plotted in Fig. 7 – even for N_{boot} as small as $N_{\text{boot}} = 5$ – this best model is correctly recovered by the AIC_p .

8.3 Is a single bootstrap iteration enough to optimize α_S ?

The small scatter in $dm_{\text{eff}}/d\alpha_S$ and $\chi^2(\alpha_S)$ that results from the high degree of correlation between models with neighbouring α_S in case of a differentiable smoothing function make the model selection with m_{eff} a very efficient ansatz to optimize the smoothing in any kind of fit. In fact, in many cases already a single bootstrap iteration can be enough to get a decent estimate of α_S .

We illustrate this in Fig. 8 that is similar to Figs 4 and 5 but the number of bootstrap iterations has been reduced to $N_{\text{boot}} = 1$ and only AIC_p , m_{eff} , and the recovery of the input model are displayed. It is remarkable how well the recovery of the input model works: after a single iteration, without any separate Monte Carlo simulations to calibrate α_S , the recovery with the smallest rms is identified.

8.4 Model correlations

It is worth looking at the differences between the non-parametric case and the parametric case. The parametric analogue to the differentiable smoothing factor α_S is the order n_{GH} of the parametric models. In

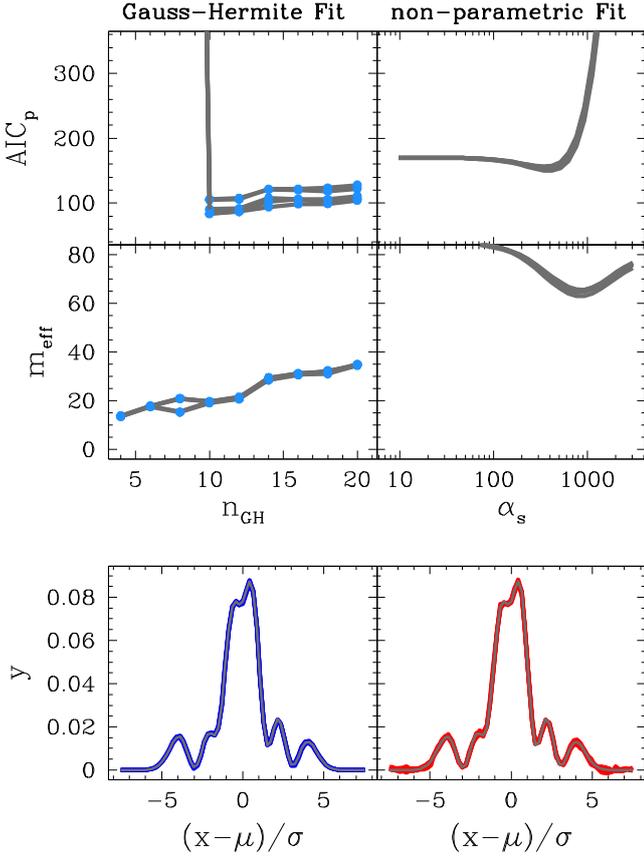


Figure 8. Similar to Figs 4 and 5 but the number of bootstrap iterations to calculate m_{eff} has been reduced to $N_{\text{boot}} = 1$. The figure only shows AIC_p and m_{eff} . The recovery of the input model is almost not affected by the highly reduced number of bootstraps.

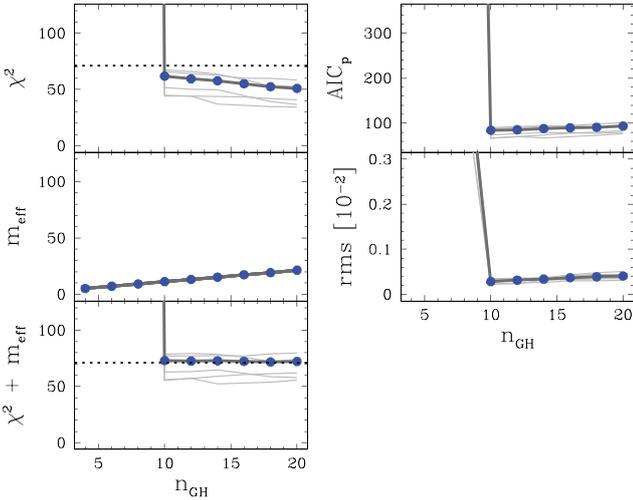


Figure 9. Left-hand panels: Same as the left-hand panels in Fig. 4, but the results have been averaged over 20 mock data sets. For comparison, the thin lines show again the results of the left-hand panels in Fig. 4. Right-hand panels: Same as the left-hand panels in Fig. 5, but the results have been averaged over 20 mock data sets. As in the left-hand panels, the results of Fig. 5 are shown for comparison as well (thin lines). This example shows how jagged χ^2 curves that result from a lack of correlation between ‘neighbouring’ models are smoothed out in the average over repeated measurements. (Such an average corresponds to the expectation E_y in equation 9.)

contrast to α_s , n_{GH} is not a differentiable parameter. Rather, it is discrete. Hence, even fits with ‘adjacent’ n_{GH} , i.e. fits at n_{GH} and $n_{\text{GH}} + 2$ are considerably different and less correlated than models with adjacent α_s are. This can be clearly seen from a comparison of the amount of scatter in the χ^2 curves shown in the top panels of Fig. 4. The discrete nature of n_{GH} in the parametric models and the respective weaker correlation between models with similar but not identical n_{GH} suppresses the covariance term in the analogue of equation (27) for the parametric case. This leads to the jagged $\chi^2(n_{\text{GH}})$ curves in the parametric case – in contrast to the smooth $\chi^2(\alpha_s)$ in the non-parametric case. When the degree of correlation between the models is low, then *both* χ^2 and m_{eff} become noisy.

As we have seen in Section 8.1, the noise in m_{eff} can be made arbitrarily small with a sufficiently large N_{boot} , i.e. by averaging over different noise patterns in the artificial *bootstrap* data. The noise in χ^2 can be reduced in an analogous way, but this requires repeated measurements, i.e. averaging over different noise patterns in the *actual* data.

This is illustrated in Fig. 9 where some of the parametric fit results that were already shown in Figs 4 and 5 are plotted again. However, in addition to the results for the five individual mock data sets, we also show the respective averages over fits to 20 mock data sets. As expected, these averages become perfectly smooth.⁵

Note that the non-parametric fits do not suffer from such a strong dependence on the mock noise since they depend in a differentiable way on the respective parameter α_s . The high degree of correlation between neighbouring models makes both the χ^2 and the m_{eff} terms well behaved in this case. The model selection via m_{eff} is therefore very efficient to obtain the optimal smoothing in non-parametric models.

9 SUMMARY

We have introduced a simple data-driven method to optimize the smoothing of parametric and non-parametric models without the need of separate Monte Carlo simulations. The method builds on a generalized concept of *effective* number of parameters (Ye 1998; Lipka & Thomas 2021) that can be easily computed for each model based on bootstrap simulations (cf. equation 7). It quantifies the complexity of a model in a very flexible way that can be used in linear as well as non-linear models, in models with or without constraint equations for the parameters, and in models with or without penalties. In the simplest situation of a penalty-free model, it reduces to the classical number of fitted parameters m .

We have shown that the concept of effective number of parameters naturally emerges when the classical ideas of model selection are extended to models that do not fulfil the maximum-likelihood condition. In this case, the classical $\text{AIC} = \chi^2 + 2m$ can be generalized to $\text{AIC}_p = \chi^2 + 2m_{\text{eff}}$. For linear models, this holds exactly. The generalized AIC_p can be applied to the large class of penalized maximum-likelihood models in particular.

As an application of the generalized model selection for penalized models, we have tested two classes of fits to some mock data loosely inspired by the problem of fitting the shape of an emission line in a galaxy spectrum. The first model class is parametric and we showed how the generalized model selection leads to the recovery of the correct order of the fitting function. In this case, the results of the generalized model selection and the classical AIC are supposed to coincide and they indeed do. In our second example, we showed how

⁵In the context of equation (9), we here perform the expectation E_y .

well the generalized model selection works to optimize the strength of a smoothing penalty in a non-parametric model. Note that for all the recoveries presented in this paper, the code was not provided a value for the smoothing factor but determined the optimal smoothing purely by itself and the data.

We have discussed in detail the efficiency of the method. While it does not require separate Monte Carlo simulations to calibrate the optimal smoothing, it requires bootstrap simulations to compute the effective number of parameters. A great advantage of the method is that for smooth model functions the number of required bootstraps is very low, of the order of 10 or even less. Each bootstrap represents a fit to a new – bootstrapped – data set, meaning that the extra cost to optimize the smoothing is to do 10 rather than 1 fit per smoothing value.

The generalized model selection has probably many astrophysical applications. In a previous paper, we have already experimented with the generalized model selection in the context of orbit superposition models with tens of thousands of parameters and non-linear smoothing (entropy) constraints. There, the best generalized model selection criterion was the one derived here and model selection turned out necessary to obtain unbiased model results (Lipka & Thomas 2021).

In a companion paper, we will introduce a new spectral fitting code that makes use of the concepts discussed here to measure non-parametric line-of-sight velocity distributions of stars in galaxies (Thomas et al., in preparation). We anticipate that the method can be applied in many other situations like non-parametric deprojections (Magorrian 1999; de Nicola et al. 2020) or any other situation where substructure has to be separated from noise like in strong gravitational lensing. Moreover, it is not tied to smoothing problems. The strength of any penalty function can be optimized in the same way as outlined here.

ACKNOWLEDGEMENT

We thank the anonymous referee for comments that helped improving the paper.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Akaike H., 1973, *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer, New York, p. 199
- Akaike H., 1974, *IEEE Trans. Autom. Control*, 19, 716
- Andrae R., Schulze-Hartung T., Melchior P., 2010, preprint ([arXiv:1012.3754](https://arxiv.org/abs/1012.3754))
- Burnham K., Anderson D., 2002, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, Berlin
- de Nicola S., Saglia R. P., Thomas J., Dehnen W., Bender R., 2020, *MNRAS*, 496, 3076
- Hastie T., Tibshirani R., Friedman J., 2013, *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Springer Science & Business Media, Berlin
- Lipka M., Thomas J., 2021, *MNRAS*, 504, 4599
- Magorrian J., 1999, *MNRAS*, 302, 530
- Myller-Lebedeff W., 1907, *Math. Ann.*, 64, 388
- Richstone D. O., Tremaine S., 1988, *ApJ*, 327, 82
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Richstone D., 2004, *MNRAS*, 353, 391
- van der Marel R. P., Franx M., 1993, *ApJ*, 407, 525
- Ye J., 1998, *J. Am. Stat. Assoc.*, 93, 120

APPENDIX A: THE EQUIVALENCE OF m AND m_{EFF} IN MODELS WITHOUT A PENALTY

The bootstrap data are constructed by adding noise to the best-fitting model obtained from the actual data. The χ^2 of the bootstrapped data with respect to the original model reads

$$\chi_{\text{prior}}^2 = \sum_{i=1}^{N_{\text{data}}} \left(\frac{z_i^k - f_i(\hat{\Theta}_y)}{\epsilon_i} \right)^2. \quad (\text{A1})$$

We call this quantity χ_{prior}^2 because it represents the residuals prior to the bootstrap fit. χ_{prior}^2 follows a χ^2 distribution with N_{data} degrees of freedom, i.e. $E(\chi_{\text{prior}}^2) = N_{\text{data}}$.

After the bootstrap fit, the χ^2 with respect to the best-fitting model obtained from the bootstrapped data reads

$$\chi_{\text{posterior}}^2 = \sum_{i=1}^{N_{\text{data}}} \left(\frac{z_i^k - f_i(\hat{\Theta}_z^k)}{\epsilon_i} \right)^2. \quad (\text{A2})$$

To contrast it from the above χ_{prior}^2 we call it $\chi_{\text{posterior}}^2$ because it represents the residuals posterior to the bootstrap fit. For a linear model with m independent variables $\chi_{\text{posterior}}^2$ follows a χ^2 distribution with $N_{\text{data}} - m$ degrees of freedom implying $E(\chi_{\text{posterior}}^2) = N_{\text{data}} - m$.

With some simple algebraic conversions, equation (7) can be written as

$$E(m_{\text{eff}}^k) = E(\chi_{\text{prior}}^2) - E(\chi_{\text{posterior}}^2) + \sum_{i=1}^{N_{\text{data}}} E \left[\left(\frac{f_i(\hat{\Theta}_z^k) - z_i^k}{\epsilon_i} \right) \left(\frac{f_i(\hat{\Theta}_z^k) - f_i(\hat{\Theta}_y)}{\epsilon_i} \right) \right] \quad (\text{A3})$$

and further transformed into

$$m_{\text{eff}} = E(\chi_{\text{prior}}^2) - E(\chi_{\text{posterior}}^2) + \sum_{i=1}^{N_{\text{data}}} E \left[\left(\frac{f_i(\hat{\Theta}_z^k) - z_i^k}{\epsilon_i^2} \right) f_i(\hat{\Theta}_z^k) \right] - \sum_{i=1}^{N_{\text{data}}} E \left[\left(\frac{f_i(\hat{\Theta}_z^k) - z_i^k}{\epsilon_i^2} \right) f_i(\hat{\Theta}_y) \right]. \quad (\text{A4})$$

In the absence of a penalty \mathcal{P} the mean of the fit is invariant under the bootstrap iterations, $E(f_i(\hat{\Theta}_z^k)) = f_i(\hat{\Theta}_y)$ (Appendix B). Since by construction $E(z_i^k) = f_i(\hat{\Theta}_y)$ the last sum of expectation values therefore vanishes. Moreover, without a penalty function it can also be shown that

$$\sum_{i=1}^{N_{\text{data}}} E \left(\frac{z_i^k f_i(\hat{\Theta}_z^k)}{\epsilon_i^2} \right) = \sum_{i=1}^{N_{\text{data}}} E \left(\frac{f_i(\hat{\Theta}_z^k) f_i(\hat{\Theta}_z^k)}{\epsilon_i^2} \right) \quad (\text{A5})$$

(Appendix B) meaning that also the first sum of expectation values in equation (A4) vanishes.

Hence, for linear models without constraints and penalties

$$m_{\text{eff}} = E(\chi_{\text{prior}}^2) - E(\chi_{\text{posterior}}^2) = m. \quad (\text{A6})$$

This means that in the absence of a penalty term, m_{eff} behaves exactly as the classical number of variables m . In this case, m is a measure of the responsiveness of the model to noise. The equality is no longer guaranteed when the parameter estimation is subject to a penalty term.

APPENDIX B: SPECIFIC PROPERTIES OF LINEAR MODELS

A linear model with m parameters can be represented by a matrix \mathbf{A} and the parameter vector $\Theta = (\Theta_1, \dots, \Theta_m)$ such that the model

vector $f = \mathbf{A}\Theta$ or, in index notation,

$$f_i = \sum_{k=1}^m A_{ik} \Theta_k. \quad (\text{B1})$$

The matrix \mathbf{A} consists of the partial derivatives of the model f with respect to the parameters Θ ,

$$\frac{\partial f_i}{\partial \Theta_k} = A_{ik}. \quad (\text{B2})$$

For simplicity, we assume uncorrelated Gaussian errors such that the log-likelihood of the model reads $\log \mathcal{L} \sim -\chi^2/2$, where

$$\chi^2 = (y - \mathbf{A}\Theta)^T \Sigma^{-1} (y - \mathbf{A}\Theta). \quad (\text{B3})$$

Here, y is the data vector and Σ is the variance–covariance matrix.

At the fitted parameter values $\hat{\Theta}_y$, the model reads

$$f_i(\hat{\Theta}_y) = \sum_{k=1}^m A_{ik} \hat{\Theta}_{y,k} \quad (\text{B4})$$

and the bootstrap can be written as

$$z_i^\kappa = \sum_{k=1}^m A_{ik} \hat{\Theta}_{y,k} + \Delta z_i^\kappa \quad (\text{B5})$$

where Δz_i^κ is the bootstrap noise of iteration κ . Finally,

$$f_i(\hat{\Theta}_z^\kappa) = \sum_k A_{ik} \hat{\Theta}_{z,k}^\kappa. \quad (\text{B6})$$

The χ^2 minimization implies

$$\sum_{i=1}^{N_{\text{data}}} \left(\frac{2(z_i^\kappa - f_i(\hat{\Theta}_z^\kappa))}{\epsilon_i^2} \frac{\partial f_i}{\partial \Theta_j}(\hat{\Theta}_z^\kappa) \right) = 0. \quad (\text{B7})$$

which for the linear models translates into the m equations

$$\sum_{i=1}^{N_{\text{data}}} \frac{2}{\epsilon_i^2} \left(\sum_{k=1}^m A_{ik} (\hat{\Theta}_{y,k} - \hat{\Theta}_{z,k}^\kappa) + \Delta z_i^\kappa \right) A_{ij} = 0 \quad (\text{B8})$$

for $j = 1, \dots, m$. Taking the expectation value of the above over many bootstrapped data samples z , we can simplify using $E(\Delta z_i^\kappa) = 0$ and with

$$B_{kj} = \sum_{i=1}^{N_{\text{data}}} A_{ki}^T A_{ij} \quad (\text{B9})$$

and $A_{ij} = A_{ij}/\epsilon_i$ these m equations read

$$\sum_{k=1}^m B_{kj} E(\hat{\Theta}_{y,k} - \hat{\Theta}_{z,k}^\kappa) = 0. \quad (\text{B10})$$

When the m variables are independent, the matrix \mathbf{A} has maximum rank and \mathbf{B} is a $m \times m$ matrix of rank m . Then, the bootstrap assumption $E(f_i(\hat{\Theta}_z^\kappa)) = f_i(\hat{\Theta}_y)$ follows because the expectation values $E(\hat{\Theta}_{y,k} - \hat{\Theta}_{z,k}^\kappa)$ have to be zero and $E(\hat{\Theta}_{z,k}^\kappa) = \hat{\Theta}_{y,k}$ implies $E(f_i(\hat{\Theta}_z^\kappa)) = f_i(\hat{\Theta}_y)$. Multiplying the m equations of the maximum-

likelihood condition

$$\sum_{i=1}^{N_{\text{data}}} \left(\frac{2(z_i^\kappa - f_i(\hat{\Theta}_z^\kappa))}{\epsilon_i^2} \frac{\partial f_i}{\partial \Theta_j}(\hat{\Theta}_z^\kappa) \right) = 0, \quad (\text{B11})$$

$j = 1, \dots, m$, each by $\hat{\Theta}_{z,j}^\kappa$

$$\sum_{i=1}^{N_{\text{data}}} \left(\frac{(z_i^\kappa - f_i(\hat{\Theta}_z^\kappa))}{\epsilon_i^2} \frac{\partial f_i}{\partial \Theta_j}(\hat{\Theta}_z^\kappa) \right) \hat{\Theta}_{z,j}^\kappa = 0, \quad (\text{B12})$$

then

$$\sum_{i=1}^{N_{\text{data}}} \sum_{j=1}^m \left(\frac{(z_i^\kappa - f_i(\hat{\Theta}_z^\kappa))}{\epsilon_i^2} \frac{\partial f_i}{\partial \Theta_j}(\hat{\Theta}_z^\kappa) \right) \hat{\Theta}_{z,j}^\kappa = 0. \quad (\text{B13})$$

and, thus, for linear models

$$\sum_{i=1}^{N_{\text{data}}} \left(\frac{z_i^\kappa - f_i(\hat{\Theta}_z^\kappa)}{\epsilon_i^2} \right) f_i(\hat{\Theta}_z^\kappa) = 0, \quad (\text{B14})$$

which means that

$$\sum_{i=1}^{N_{\text{data}}} E \left(\frac{z_i^\kappa f_i(\hat{\Theta}_z^\kappa)}{\epsilon_i^2} \right) = \sum_{i=1}^{N_{\text{data}}} E \left(\frac{f_i(\hat{\Theta}_z^\kappa) f_i(\hat{\Theta}_z^\kappa)}{\epsilon_i^2} \right). \quad (\text{B15})$$

Under the condition of a penalized maximum-likelihood, the penalty function \mathcal{P} modifies equation (B7) to

$$\sum_{i=1}^{N_{\text{data}}} \left(\frac{2(z_i^\kappa - f_i(\hat{\Theta}_z^\kappa))}{\epsilon_i^2} \frac{\partial f_i}{\partial \Theta_j}(\hat{\Theta}_z^\kappa) \right) - \alpha_S \frac{\partial \mathcal{P}}{\partial \Theta_j} = 0. \quad (\text{B16})$$

Consequently, even for linear models neither $E(f_i(\hat{\Theta}_z^\kappa)) = f_i(\hat{\Theta}_y)$ nor equation (B15) can be assumed to hold in this case.

For linear models as above

$$\frac{\partial \log \mathcal{L}(y|\Theta)}{\partial \Theta} = (y - \mathbf{A}\Theta)^T \Sigma^{-1} \mathbf{A} \quad (\text{B17})$$

and

$$\frac{\partial^2 \log \mathcal{L}(y|\Theta)}{\partial \Theta^2} = -\mathbf{A}^T \Sigma^{-1} \mathbf{A}. \quad (\text{B18})$$

For some parameter vector $\tilde{\Theta}$, this implies

$$\left(\frac{\partial \log \mathcal{L}(y|\Theta)}{\partial \Theta} \right) \tilde{\Theta} = (y - f(\tilde{\Theta}))^T \Sigma^{-1} f(\tilde{\Theta}) \quad (\text{B19})$$

and

$$\tilde{\Theta}^T \left(\frac{\partial^2 \log \mathcal{L}(y|\Theta)}{\partial \Theta^2} \right) \tilde{\Theta} = -f(\tilde{\Theta})^T \Sigma^{-1} f(\tilde{\Theta}). \quad (\text{B20})$$

In particular, the last expression does not depend on Θ but only on $\tilde{\Theta}$.

For linear models, the definition of m_{eff} (cf. equation 6) can be written as

$$m_{\text{eff}} = E_z \left[(f(\hat{\Theta}_z) - f(\hat{\Theta}_y))^T \Sigma^{-1} (z - f(\hat{\Theta}_y)) \right]. \quad (\text{B21})$$

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

Bibliography

- Akaike H., 1973, *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer New York, New York, NY, pp 199–213
- Akaike H., 1974, *IEEE Transactions on Automatic Control*, 19, 716
- Andrae R., Schulze-Hartung T., Melchior P., 2010, arXiv e-prints, p. arXiv:1012.3754
- Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Bundy K., Moustakas L. A., 2009, *ApJ*, 705, 1099
- Baade W., 1944, *ApJ*, 100, 147
- Barnes J. E., Hernquist L., 1992, *Nature*, 360, 715
- Bender R., Moellenhoff C., 1987, *A&A*, 177, 71
- Bender R., Doebereiner S., Moellenhoff C., 1988, *A&AS*, 74, 385
- Bender R., Paquet A., Nieto J. L., 1991, *A&A*, 246, 349
- Bender R., Saglia R. P., Gerhard O. E., 1994, *MNRAS*, 269, 785
- Binggeli B., Jerjen H., 1998, *A&A*, 333, 17
- Binggeli B., Tammann G. A., Sandage A., 1987, *AJ*, 94, 251
- Binney J., Mamon G. A., 1982, *MNRAS*, 200, 361
- Binney J., Tremaine S., 2008, *Galactic Dynamics: Second Edition*. Princeton University Press
- Blanton M. R., et al., 2017, *AJ*, 154, 28
- Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, *ApJ*, 539, 517
- Burnham K., Anderson D., 2002, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag
- Cappellari M., 2002, *MNRAS*, 333, 400

- Cappellari M., 2008, MNRAS, 390, 71
- Cappellari M., 2023, MNRAS, 526, 3273
- Cappellari M., Copin Y., 2003, MNRAS, 342, 345
- Cappellari M., et al., 2011, MNRAS, 413, 813
- Cloet-Osselaer A., De Rijcke S., Vandenbroucke B., Schroyen J., Koleva M., Verbeke R., 2014, MNRAS, 442, 2909
- Contopoulos G., 1963, AJ, 68, 1
- Dehnen W., Gerhard O. E., 1993, MNRAS, 261, 311
- Dekel A., Silk J., 1986, ApJ, 303, 39
- Djorgovski S., Davis M., 1987, ApJ, 313, 59
- Dressler A., 1980, ApJ, 236, 351
- Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, ApJ, 313, 42
- Erwin P., 2018, MNRAS, 474, 5372
- Faber S. M., et al., 1997, AJ, 114, 1771
- Fabricius M. H., Barnes S., Bender R., Drory N., Grupp F., Hill G. J., Hopp U., MacQueen P. J., 2008, in McLean I. S., Casali M. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II. p. 701473, doi:10.1117/12.787204
- Fabricius M. H., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84465K, doi:10.1117/12.925177
- Falcón-Barroso J., Martíg M., 2021, A&A, 646, A31
- Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, A&A, 532, A95
- Ferguson H. C., Binggeli B., 1994, A&ARv, 6, 67
- Gallagher John S. I., Wyse R. F. G., 1994, PASP, 106, 1225
- Gebhardt K., et al., 2001, AJ, 122, 2469

- Gebhardt K., et al., 2003, *ApJ*, 583, 92
- Geha M., Guhathakurta P., van der Marel R. P., 2003, *AJ*, 126, 1794
- Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, *ApJ*, 757, 85
- Gerhard O. E., 1993, *MNRAS*, 265, 213
- Gerhard O. E., Binney J. J., 1996, *MNRAS*, 279, 993
- Gieles M., Zocchi A., 2015, *MNRAS*, 454, 576
- Graham A. W., Erwin P., Trujillo I., Asensio Ramos A., 2003, *AJ*, 125, 2951
- Gunn J. E., Gott J. Richard I., 1972, *ApJ*, 176, 1
- Hernquist L., 1990, *ApJ*, 356, 359
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, *ApJS*, 175, 356
- Hubble E. P., 1927, *The Observatory*, 50, 276
- Hubble E. P., 1936, *Realm of the Nebulae*
- Janz J., et al., 2014, *ApJ*, 786, 105
- Jeans J. H., 1915, *MNRAS*, 76, 70
- Jeans J. H., 1922, *MNRAS*, 82, 122
- Jerjen H., Kalnajs A., Binggeli B., 2000, *A&A*, 358, 845
- King I. R., 1966, *AJ*, 71, 64
- Kluge M., et al., 2020, *ApJS*, 247, 43
- Kormendy J., 1977, *ApJ*, 218, 333
- Kormendy J., Bender R., 1996, *ApJ*, 464, L119
- Kormendy J., Bender R., 2012, *ApJS*, 198, 2
- Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, *ApJS*, 182, 216
- Larson R. B., Tinsley B. M., Caldwell C. N., 1980, *ApJ*, 237, 692
- Liepold C. M., Quenneville M. E., Ma C.-P., Walsh J. L., McConnell N. J., Greene J. E., Blakeslee J. P., 2020, *ApJ*, 891, 4
- Lipka M., Thomas J., 2021, *MNRAS*, 504, 4599

- Lisker T., Grebel E. K., Binggeli B., 2006a, *AJ*, 132, 497
- Lisker T., Glatt K., Westera P., Grebel E. K., 2006b, *AJ*, 132, 2432
- Loebman S. R., Ivezić Ž., Quinn T. R., Governato F., Brooks A. M., Christensen C. R., Jurić M., 2012, *ApJ*, 758, L23
- Lokas E. L., 2002, *MNRAS*, 333, 697
- Long R. J., Mao S., 2010, *MNRAS*, 405, 301
- Magorrian J., 1999, *MNRAS*, 302, 530
- Magorrian J., Binney J., 1994, *MNRAS*, 271, 949
- Mehrgan K., Thomas J., Saglia R., Parikh T., Bender R., 2023, *ApJ*, 948, 79
- Mehrgan K., Thomas J., Saglia R., Parikh T., Neureiter B., Erwin P., Bender R., 2024, *ApJ*, 961, 127
- Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, *ApJ*, 524, L19
- Napolitano N. R., Pota V., Romanowsky A. J., Forbes D. A., Brodie J. P., Foster C., 2014, *MNRAS*, 439, 659
- Neureiter B., et al., 2021, *MNRAS*, 500, 1437
- Neureiter B., Thomas J., Rantala A., Naab T., Mehrgan K., Saglia R., de Nicola S., Bender R., 2023, *ApJ*, 950, 15
- Nowak N., Thomas J., Erwin P., Saglia R. P., Bender R., Davies R. I., 2010, *MNRAS*, 403, 646
- Onken C. A., et al., 2007, *ApJ*, 670, 105
- Parikh T., et al., 2018, *MNRAS*, 477, 3954
- Penny S. J., Forbes D. A., Pimblet K. A., Floyd D. J. E., 2014, *MNRAS*, 443, 3381
- Prugniel P., Soubiran C., Koleva M., Le Borgne D., 2007, arXiv e-prints, pp astro-ph/0703658
- Quenneville M. E., Liepold C. M., Ma C.-P., 2022, *ApJ*, 926, 30
- Read J. I., Steger P., 2017, *MNRAS*, 471, 4541
- Richstone D. O., Tremaine S., 1988, *ApJ*, 327, 82

- Romero-Gómez J., Aguerri J. A. L., Peletier R. F., Mieske S., van de Ven G., Falcón-Barroso J., 2024, *MNRAS*, 527, 9715
- Rusli S. P., et al., 2013, *AJ*, 146, 45
- Rybicki G. B., 1987, in de Zeeuw P. T., ed., *IAU Symposium Vol. 127, Structure and Dynamics of Elliptical Galaxies*. p. 397, doi:10.1007/978-94-009-3971-4_41
- Sandage A., Binggeli B., Tammann G. A., 1985, *AJ*, 90, 1759
- Schechter P., 1976, *ApJ*, 203, 297
- Schwarz G., 1978, *The Annals of Statistics*, 6, 461
- Schwarzschild M., 1979, *ApJ*, 232, 236
- Seo M., Ann H. B., 2022, *MNRAS*, 514, 5853
- Seo M., Ann H. B., 2023, *MNRAS*, 520, 5521
- Sérsic J. L., 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 6, 41
- Simien F., Prugniel P., 2002, *A&A*, 384, 371
- Siopis C., et al., 2009, *ApJ*, 693, 946
- Strigari L. E., Bullock J. S., Kaplinghat M., 2007, *ApJ*, 657, L1
- Syer D., Tremaine S., 1996, *MNRAS*, 282, 223
- Thater S., et al., 2022, *MNRAS*, 509, 5416
- Thomas J., Lipka M., 2022, *MNRAS*, 514, 6203
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Richstone D., 2004, *MNRAS*, 353, 391
- Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2007, *MNRAS*, 382, 657
- Thomas J., et al., 2011, *MNRAS*, 415, 545
- Toomre A., 1982, *ApJ*, 259, 535
- Walker M. G., Peñarrubia J., 2011, *ApJ*, 742, 20
- Watkins L. L., van de Ven G., den Brok M., van den Bosch R. C. E., 2013, *MNRAS*, 436, 2598

Weisz D. R., et al., 2011, *ApJ*, 739, 5

Wilson C. P., 1975, *AJ*, 80, 175

Zöller R., Kluge M., Staiger B., Bender R., 2024, *ApJS*, 271, 52

de Lorenzi F., Debattista V. P., Gerhard O., Sambhus N., 2007, *MNRAS*, 376, 71

de Nicola S., Saglia R. P., Thomas J., Dehnen W., Bender R., 2020, *MNRAS*, 496, 3076

de Vaucouleurs G., 1948, *Annales d'Astrophysique*, 11, 247

de Vaucouleurs G., 1961, *ApJS*, 5, 233

den Brok M., Krajnović D., Emsellem E., Brinchmann J., Maseda M., 2021, *MNRAS*, 508, 4786

van Zee L., Skillman E. D., Haynes M. P., 2004, *AJ*, 128, 121

van den Bergh S., 1976, *ApJ*, 206, 883

van den Bergh S., 2000, *The Galaxies of the Local Group*

van der Marel R. P., Franx M., 1993, *ApJ*, 407, 525

Acknowledgements

I want to thank PD Dr. Roberto Saglia for providing me the opportunity and continuous support to write this PhD thesis. I also want to express my deepest gratitude to Dr. Jens Thomas, the extensive discussions, help, and counsel that was provided by You proofed invaluable to this work. Thank You for always having time for me when I had questions and inspiring me to dig deeper. I was able to learn a lot from your expertise, both in physics and outside of it. I would also like to thank all of my co-authors that made the VIRUS-W project possible in the first place. Last, but not least, I am deeply grateful to my friends and family. Thank you to my parents, Renate and Werner, and my sister, Barbara, for your unwavering support and for always being there for me when I needed you.