
Accelerated Bayesian inference techniques for event based datasets in nested sampling and X-ray astronomy

Margret Westerkamp



München 2025

Accelerated Bayesian inference techniques for event based datasets in nested sampling and X-ray astronomy

Margret Westerkamp

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

vorgelegt von
Margret Westerkamp
aus Langenfeld

München, den 12.12.2024

Erstgutachter: PD Dr. Torsten Enßlin
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 06.03.2025

Abstract

The updating of our knowledge in response to new information can prove to be complex, both in everyday life and in science. Bayes' theorem provides a systematic, probabilistic framework for updating our knowledge given new data. In any case of Bayesian inference - parameter estimation, model comparison, or field inference - probability distributions are used to encode our knowledge about the quantity of interest. This quantity is referred to as the signal. To compute the posterior probability of the signal in the presence of new data, Bayes' theorem combines prior knowledge of the signal with knowledge of the data measurement process, represented by the likelihood. This probabilistic approach not only provides a posterior estimate of the signal, but also allows for uncertainty quantification.

This thesis focuses on Bayesian field inference built upon information field theory. Inferring a field from an inherently finite data set is an under-constrained problem. Hence, the inclusion of prior knowledge is essential. In the present work, the implementation of generative, non-parametric prior models allows to exploit possibly complex and a priori unknown correlation structures of the signal during inference. Two distinct applications of Bayesian field inference are discussed: Bayesian evidence calculation and imaging.

The methodological part addresses the calculation of the posterior normalization - the Bayesian evidence. The evidence plays an important role in Bayesian model comparison. However, the evidence may be computationally intractable, for example due to complex relationships between the prior and posterior. Nested sampling provides a numerical estimate of the evidence by examining the likelihood as a function of enclosed prior volumes. In particular, the algorithm is based on statistical estimates of the prior volumes, introducing a stochastic error. For this reason a one-dimensional Bayesian field inference problem is formulated to obtain improved estimates of the prior volumes and the corresponding evidence.

The second and main part is dedicated to advancing Bayesian imaging of the X-ray sky. Observed by space-based telescopes, X-rays allow us to study some of the most energetic phenomena in the universe. However, the interpretability of the data is limited by overlapping X-ray sources, Poisson noise, and instrumental effects such as the point spread function. In this context, Bayesian forward models are constructed for the X-ray telescopes Chandra and eROSITA, tailored to their instrumental properties, to obtain denoised, deconvolved, decomposed, and spatio-spectral resolved images of the X-ray object of interest. The work presented includes the development of the JAX-accelerated open source software package J-UBIK to support future and existing Bayesian imaging models. Imaging results for the supernova remnant of SN1006 and for the Large Magellanic Cloud provide a detailed view of the diffuse X-ray structures and the separated point sources. This paves the way for future analysis of such fine-scale structures, such as shock fronts of supernova remnants, the construction of point source catalogs, and possibly improved instrument calibration.

Zusammenfassung

Unser Wissen an neue Informationen anzupassen, kann sowohl im Alltag als auch in der Wissenschaft eine Herausforderung darstellen. Der Satz von Bayes bietet einen systematischen, probabilistischen Ansatz zur Aktualisierung unseres Wissens auf der Grundlage neuer Daten. Bei der Bayes'schen Inferenz, sei es bei der Parameterschätzung, dem Modellvergleich oder der Feldinferenz, werden Wahrscheinlichkeitsverteilungen als Repräsentation unseres Wissens über die hier relevante Größe, das Signal, verwendet. Unter Verwendung von Vorwissen, auch A-priori-Wissen, über das Signal und der so genannten Likelihood, die den Messprozess der Daten beschreibt, ermöglicht der Satz von Bayes, die A-posteriori-Wahrscheinlichkeit des Signals unter Berücksichtigung der Daten zu inferieren. Somit liefert der probabilistische Ansatz nicht nur eine A-posteriori-Schätzung des Signals, sondern ermöglicht auch die Quantifizierung der entsprechenden Unsicherheit.

Der Schwerpunkt dieser Dissertation liegt auf der Bayes'schen Feldinferenz basierend auf der Informationsfeldtheorie. Die Inferenz eines Feldes aus einem inhärent endlichen Datensatz ist ein unterbestimmtes Problem. Daher ist die Einbeziehung von A-priori-Wissen unerlässlich. Die Implementierung von generativen, nichtparametrischen Prior-Modellen ermöglicht es, potentiell komplexe und a priori unbekannte Korrelationsstrukturen des Signals in die Inferenz mit einzubeziehen. Insgesamt umfasst diese Arbeit zwei verschiedene Anwendungen der Bayes'schen Inferenz: Bayes'sche Evidenzberechnung und Bildgebung.

Die Berechnung der Normalisierung des Posteriors - der Bayes'schen Evidenz - ist Gegenstand des methodischen Teils. Die Evidenz spielt eine wichtige Rolle beim Vergleich Bayes'scher Modelle. Allerdings können unter anderem komplexe Beziehungen zwischen dem Prior und dem Posterior dazu führen, dass die Evidenz nicht analytisch berechnet werden kann. Ein numerisches Verfahren, das als "Nested Sampling" bezeichnet wird, liefert eine Schätzung der Evidenz, indem sie die Likelihood als Funktion eingeschlossener Prior-Volumina sondiert. Präziser formuliert basiert der Algorithmus auf einer statistischen Schätzung der Prior-Volumina, die einen stochastischen Fehler in die Evidenzberechnung einführt. Um verbesserte Schätzungen der Prior-Volumina und der entsprechenden Evidenz zu erhalten, wird ein eindimensionales Bayes'sches Feldinferenzproblem formuliert.

Der zweite Teil befasst sich mit der Bayes'schen Bildgebung des Röntgenhimmels. Durch die Beobachtung von Röntgenstrahlen mit Teleskopen außerhalb unserer Atmosphäre können wir einige der energiereichsten Phänomene im Universum untersuchen. Die Interpretierbarkeit der Daten wird jedoch durch überlappende Röntgenquellen, Poisson-Rauschen und instrumentelle Effekte wie die Punktspreizfunktion eingeschränkt. Hier werden Bayes'sche Vorwärtsmodelle für die Röntgenteleskope Chandra und eROSITA erstellt, die auf ihre instrumentellen Eigenschaften zugeschnitten sind, um entrauschte, dekonvolvierte, zerlegte und räumlich-spektral aufgelöste Bilder des betreffenden Röntgenobjekts zu erhalten. Zur Unterstützung zukünftiger und bestehender Bayes'scher Bildgebungsmodelle wird das JAX-beschleunigte Open-Source-Softwarepaket J-UBIK entwickelt. Die Bildgebungsergebnisse für den Supernovaüberrest von SN1006 und für die Große Magellansche Wolke bieten einen detaillierten Blick auf die diffusen Röntgenfeinstrukturen und die davon separierten Punktquellen. Dies ebnet den Weg für die zukünftige Analyse solcher Strukturen, wie z. B. Schockfronten von Supernova-Überresten, die Erstellung von Punktquellenkatalogen und möglicherweise eine verbesserte Instrumentenkalibrierung.

Acknowledgement

I would like to extend my heartfelt gratitude to everyone who inspired, motivated, and played a crucial role in the success of this work.

My deepest gratitude goes to my supervisor, *Torsten Enßlin*, for his unwavering support throughout and before my PhD journey. His constructive feedback, whether on drafts, scientific results or presentations, has not only significantly shaped my thesis, but also enriched my personal and scientific development. Discussions with him and his vision have always been a great inspiration for me. I really appreciate that he always took the time to discuss questions and concerns in detail and I consider myself very lucky to have been part of Torsten's IFT group.

I would also like to express my gratitude to *Philipp Frank* for his invaluable scientific input, constructive feedback, and the many insightful blackboard discussions. His statistical knowledge was crucial for the validation and analysis of the results presented in the thesis. I am very grateful for all that I have learned from our discussions.

I would like to especially acknowledge the collaboration with *Vincent Eberle* and *Matteo Guardiani*. Thank you, Vincent, for helping me get started in the early stages and for proofreading the final draft of my thesis. Your willingness to help, your precision, and your sense of humor have left a lasting impression on me — my time at the MPA would not have been the same without you. Thank you, Matteo, for being a wonderful office neighbor and for bringing your Italian flair and outlook on life to our daily interactions. Your curiosity and thought-provoking questions often challenged me to think more deeply and encouraged me to question things more myself. I am deeply grateful to both of you for the countless discussions, the joint sprints, the time spent together at conferences, and the ever-growing art gallery we shared. It was truly a pleasure working with you!

I really appreciate the insightful discussions with *Jakob Roth* about nested sampling. Working together to explore new approaches was both enjoyable and rewarding, and his insights into the Gaussian approximation from lensing were a great contribution to the project. A big thank you as well to *Gordian Edenhofer*, Jakob Roth, and Philipp Frank for advancing the development of NIFTy.re, which formed the foundation for the J-UBIK presented in this work.

During my time in the IFT group, the numerous discussions with both group members and those associated with the group — whether in seminars, over coffee, or at the Schneefernerhaus — were invaluable to me. Among many and the mentioned above, I would like to particularly thank *Johannes Harth-Kitzerow*, *Vishal Johnson*, *Viktoria Kainz*,

Hanieh Zandinejad, Richard Fuchs, Julian Rüstig, and Martin Reinecke for the fantastic time and enriching conversations.

I would also like to thank all former members of the IFT group, with whom I had the pleasure of collaborating during my master’s thesis. In particular, I am grateful for the discussions during my PhD with *Philipp Arras* and *Jakob Knollmüller*, whose knowledge and experience were always incredibly helpful and inspiring, sparking new ideas and approaches throughout my work.

I would also like to thank *Will Handley* for the enriching discussions on nested sampling and for introducing me to the `anesthetic` software. In this context, I am also grateful to MIAPbP for providing us with the opportunity and platform to develop and discuss the project. A big thank you goes to *Julia Stadler* for building the Chandra pipeline and for her prompt and constructive feedback on paper drafts. My gratitude also extends to *Lukas Scheel-Platz*, whose insights on the gamma-ray multi-frequency model were extremely helpful. Furthermore, I would like to thank *Mara Salvato* and *Michael Freyberg* for their support and knowledge-sharing within the eROSITA project, as well as *Selma de Mink* and *Marat Gilfanov* for being part of my PhD committee at the MPA and providing valuable input throughout my PhD journey.

I would like to sincerely thank the DLR for providing the funding, which made it possible for us to bring the UBIK project to life.

Last but not least, I want to thank those closest to me. To my husband, thank you for your patience, your loving support and the final proofreading. I can hardly put into words how grateful I am for your assistance during the final stages, which allowed me to focus fully on my work. To my parents, your constant support and love have been my foundation throughout this whole journey. Knowing you are always there for me means the world to me. To my brother, thank you for always for lending a listening ear, offering thoughtful feedback, and helping with proofreading. I’m grateful that I can always count on you. And to all my friends, thank you for your encouragement and genuine interest in my work. I feel incredibly fortunate to have a social circle that provides so much strength and motivation.

¹I would also like to acknowledge the use of ChatGPT and DeepL. I used DeepL for translation, improving phrasing, and checking spelling and grammar. ChatGPT was helpful for finding synonyms and alternative expressions, allowing me to refine the language and enhance clarity.

Contents

Abstract	iii
Zusammenfassung	v
Acknowledgement	vii
1 Introduction	1
1.1 Bayesian imaging of physical scalar fields	2
1.1.1 Probability models	4
1.1.2 Posterior approximation	6
1.1.3 Model evaluation	8
1.2 X-ray astronomy and its observations	8
1.2.1 X-ray emitting astronomical objects and there emission processes .	9
1.2.2 X-ray telescopes	10
1.3 Bayesian X-ray imaging	12
1.4 Objectives and structure of the thesis	13
1.5 Additional work	15
2 Field based Bayesian evidence inference from nested sampling data	17
2.1 Introduction	18
2.2 Methods	21
2.2.1 Nested sampling algorithm	21
2.2.2 Gaussian processes and information field theory	25
2.2.3 Bayesian inference of the likelihood–prior–volume function	27
2.3 Results and analysis	31
2.3.1 Gaussian case	31
2.3.2 Spike-and-slab case	32
2.4 Discussion	35
2.5 Acknowledgements	38
3 First spatio-spectral Bayesian imaging of SN1006 in X-rays	39
3.1 Introduction	40
3.2 Related work	43

3.2.1	State-of-the-art X-ray imaging	43
3.2.2	Previous studies of SN1006 in the X-ray range	45
3.2.3	Previous work on high energy count data with IFT	46
3.3	Image reconstruction with IFT	46
3.4	Algorithm overview of Bayesian inference of the X-ray sky	48
3.4.1	Structure of the reconstruction algorithm	48
3.4.2	Variational inference and generative models	49
3.4.3	Transition model	49
3.5	Prior models for the X-ray sky	50
3.5.1	Prior composition	50
3.5.2	Correlated components	52
3.5.3	Point-like components	53
3.6	Chandra instrument and data description	54
3.6.1	Chandra instrument response	55
3.6.2	Chandra data of SN1006	57
3.7	Validation of the algorithm using synthetic data	58
3.8	Results and analysis of inference performance	61
3.8.1	Quantitative discussion	61
3.8.2	Analysis and comparison with previous studies	64
3.9	Conclusion	67
3.10	Acknowledgements	68
4	Bayesian multi-wavelength imaging of the eROSITA LMC SN1987A	69
4.1	Introduction	70
4.1.1	Related work	70
4.2	Observations	72
4.3	Mathematical derivations	73
4.3.1	Imaging with information field theory	73
4.3.2	Prior models	74
4.3.3	The likelihood	76
4.3.4	Inference	80
4.4	Results	81
4.5	Discussion and validation	83
4.5.1	Validation	83
4.5.2	Discussion of results	85
4.6	Conclusion	86
5	J-UBIK: The JAX-accelerated Universal Bayesian Imaging Kit	87
5.1	Summary	88
5.2	Statement of need	89
5.3	Bayesian imaging with J-UBIK	90
5.3.1	Prior models	90
5.3.2	Likelihood models	91

6	Conclusion	93
A	Supplementary material chapter 2	97
A.1	Inference with Gaussian approximation	97
A.2	Reparametrization	98
A.3	Maximal likelihood calculation	99
B	Supplementary material chapter 3	101
B.1	Synthetic data generation	101
B.2	Computational analysis	101
B.3	Further diagnostics for synthetic data reconstruction	105
B.4	Further diagnostics for SN1006 reconstruction	107
C	Supplementary material chapter 4	110
C.1	eROSITA data:	
	LMC SN1987A	111
C.2	Hyper-parameter search	112
C.3	Results diagnostics	114
C.4	Validation diagnostics	118

List of Figures

1.1	General structure of a Wolter type I telescope. Incoming X-rays from the right are reflected first by the parabolic reflector (yellow) and then by the hyperbolic reflector (blue) to form an image.	11
1.2	Core components of UBIK- The Universal Bayesian Imaging Kit.	14
2.1	Illustration of the nested sampling (NS) output for simple Gaussian example introduced in Skilling (2006) and further elaborated in section 2.3.1 with two live points. Figure 2.1a shows the full NS data generated with <code>anesthetic</code> and fig. 2.1b shows a zoomed-in section, which is indicated in fig. 2.1a. The zoomed-in image additionally shows the information of the likelihood dead contours d_L , which we use as data for the Bayesian inference of the prior volumes. In both figures, the samples of likelihood–prior–volume functions defined by prior volume samples, X_k , $d_L(X_k)$, are shown as well as a likelihood–prior–volume function defined by the deterministic NS approach in eq. (2.8), $d_L(\bar{X})$	23
2.2	Visualization of the effect of changes in the mean value for one of the hyper-parameters of the power spectrum model. The changed amplitude spectrum is shown on the left side and the according influence on the field realization is shown on the right side by comparison of a reference correlated field (ref) and a variation of one of its hyper-parameters (var). Which hyper-parameter is changed and how it is changed in comparison to the reference is denoted in table 2.1.	28

2.3	Illustration of the Bayesian field inference process for the simple Gaussian, which is further elaborated in section 2.3.1 for two live points. The prior samples, the data used, and the final reconstruction compared to the NS approach and the ground truth are shown. Figure 2.3a: Prior samples for the likelihood–prior–volume function ($L^*(X)$) in yellow together with the ground truth. A zoomed-in area is marked beside it (the same area as in fig. 2.1a) which is taken to zoom into the data in fig. 2.3b and the reconstruction in fig. 2.3d. Figure 2.3b: Data on likelihood dead contours for the given zoom area. Figures 2.3c, 2.3d: Reconstruction mean (rec mean) of the likelihood–prior–volume function and the associated uncertainty, defined via the one sigma contours (rec uncertainty), which is zoomed in on in fig. 2.3d, and the full image is presented in fig. 2.3c. The result for the likelihood–prior–volume function for the deterministic NS approach is shown ($d_L(\bar{X})$), which is the same as in fig. 2.1b, and the analytic likelihood–prior–volume-function (ground truth).	30
2.4	Reconstruction results for the Gaussian prior volumes accompanied by the likelihood contours given by NS on the left and the computed log-evidence on the right for $n_{\text{live}} \in \{2, 10, 1000\}$ from top to bottom. The inferred posterior samples (rec samples) are shown together with their mean (rec mean) and compared with the corresponding statistical (NS samples) and deterministic (det NS mean) NS results and the ground truth.	33
2.5	Reconstruction results for the spike-and-slab prior volumes accompanied by the likelihood contours given by NS on the left and the computed log-evidence on the right for $n_{\text{live}} \in \{2, 10, 1000\}$ from top to bottom. The inferred posterior samples (rec samples) are shown together with their mean (rec mean) and compared with the corresponding statistical (NS samples) and deterministic (det NS mean) NS results and the ground truth.	36
3.1	Structure of the transition model given the generative prior models for the single frequency (SF) reconstruction $s_{\text{SF}} = f_{\text{SF}}(\xi_{\text{SF}})$ and for the multifrequency (MF) reconstruction $s_{\text{MF}} = f_{\text{MF}}(\xi_{\text{MF}})$, which transform the according latent parameters ξ_{SF} and ξ_{MF} from the latent space into the signal space. Here, $\xi_{\text{I, MF}}$ denotes the initial position of the MF reconstruction in latent space.	51
3.2	Visualization of the exposures used for the reconstruction in [s cm ²](table 3.3): Top: Full exposure for all front-illuminated (FI) chips. Bottom: Full exposure for all back-illuminated (BI) chips.	56
3.3	Visualization of the photon count data used for the reconstruction (table 3.3) with right ascension on the x-axis and declination on the y-axis: red = 0.5 - 1.2 keV, green = 1.2 - 2.0 keV, and blue = 2.0 - 7.0 keV.	57

3.4	Generation of synthetic data: Left: Sky sample generated for the validation experiment. Center: Chandra exposure, modeled by combining three patches (14423, 14424, 14435) (see table 3.3). Right: Synthetic data corresponding to the sky sample, obtained by convolving the sky sample with the point spread function (PSF) and drawing a pixel-wise Poisson sample from the resulting detector flux prediction.	59
3.5	Reconstruction results on synthetic data: Left: Sky sample generated for this study masked by the extent of the exposure patches (14423, 14424, 14435) (see table 3.3). Center: Reconstruction result, i.e., the posterior mean, of the imaged sky masked by the extent of the same exposure patches. Right: Zoomed-in regions of the data on top and of the reconstructed image below. The shown cutout region is marked in the center image.	59
3.6	Visualization of the two-dimensional (2D) histogram for the sample averaged relative distance of the posterior sky flux samples vs. the ground truth sky flux (eq. (3.21)). The detection limit is determined via the intersection of the line showing the mean standardized error $\bar{\epsilon}_{\text{rel}}$ (eq. (3.22)) with the $\epsilon_{\text{rel}} = 1$ line.	60
3.7	Spatial reconstruction result for the highest energy bin in $[\text{s}^{-1} \text{cm}^{-2}]$	61
3.8	Reconstruction results for the flux in $[\text{s}^{-1} \text{cm}^{-2}]$ (red = 0.5-1.2 keV, green = 1.2-2.0 keV, and blue = 2.0-7.0 keV. The corresponding color bars for each of these energy bins can be found in fig. B.9): Top: Full sky reconstruction mean. Bottom left: Reconstruction mean for diffuse emission. Bottom right: Reconstruction mean for point sources.	62
3.9	northeast (NE) quarter of SN1006 and its components: From left to right: Total sky, diffuse emission, point sources, and BI background.	63
3.10	Exposures that capture the not well separated point source (marked red). .	64
3.11	Zoom-in on the reconstruction results in the diffuse component: (a) the whole diffuse reconstruction and the locations of the zoomed-in areas; (b) the top panels represent the green areas marked in the remnant and show zoom-ins on the denoised shell of the remnant. The lower panels are represented by the white boxes in (a) and show structures in the inner soft X-ray emission of the remnant.	65
3.12	Flux intensity profiles in $[\text{s}^{-1} \text{cm}^{-2}]$: The center image shows the location of the lines along which we present the intensity profile in pixel coordinates. The corresponding intensity profiles are plotted next to the line. The posterior mean of the reconstructed flux is plotted in red and the corresponding posterior samples are plotted in grey. The profiles are shown from left to right from the outsides to the insides of the supernova remnant (SNR). . .	66
4.1	Visualisation of the exposure corrected eROSITA EDR TM8 data of ObsID 700161 for three energy bins. Red: 0.2 - 1.0 keV, Green: 1.0 - 2.0 keV, Blue: 2.0 - 4.5 keV. The white box marks the region of 30 Doradus C further discussed in section 4.3.2.	72

4.2	Visualization of one prior sample drawn from the prior model described in section 4.3.2 for three energy bins in $[1/(\text{arcsec}^2 \times \text{s})]$	76
4.3	Visualisation of the computational graph of the forward model.	80
4.4	Posterior mean of the SN1987A reconstruction. <i>The top panels</i> display on the left the reconstruction of the sky and on the right the separated diffuse emission. <i>The bottom panels</i> display the reconstruction of the point-like emission (left) and the reconstruction of the diffuse emission from 30 Doradus C (right). We convolve the point sources with an unnormalized Gaussian kernel with standard deviation $\sigma = 0.5$, in order to make them visible on printed paper. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively.	82
4.5	Visualisation of the simulated X-ray, the corresponding simulated data and its reconstruction. Left: Simulated X-ray sky. Center: Simulated X-ray data generated as shown in fig. 4.3. Right: Reconstructed X-ray sky. . . .	83
4.6	Two-dimensional histogram of the standardized error eq. (4.22) for the point sources. The histogram is plotted together with the detection threshold, $\theta = 2.5 \times 10^{-9}$, calculated in appendix C.2.	84
4.7	Zoom-in on reconstruction of diffuse emission from the Tarantula Nebula. From left to right showing the zoom area on the plot of the eROSITA LMC data, the zoomed LMC data for telescope module (TM)1, the corresponding single-TM reconstruction for TM1, the zoomed data for all 5 TMs and the reconstruction by means of all five observations.	85
5.1	Simulated X-ray sky	90
5.2	Simulated X-ray data	91
B.1	Illustration of generation of simulated data for three prior samples, showing the variance in intensity and correlation structure permitted by the prior. .	101
B.2	Time complexity (top-left: 64×64 spatial pixels; top-right: 128×128 spatial pixels; bottom-left: 256×256 spatial pixels; bottom-right: 512×512 spatial pixels). The time complexity is plotted for the different models. In green, we show just the MF reconstruction times per iteration. Light blue: duration for each iteration in the SF model before the transition. Dark blue: duration of the MF model iterations after the transition. The first dark blue marker also includes the transition time.	102

B.3	Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged sky. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.	103
B.4	Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged point source component. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.	104
B.5	Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged diffuse component. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.	105
B.6	Synthetic data reconstruction uncertainty weighted residuals (UWRs) (top row) and residuals (bottom row) for the individual energy bins (left: 0.5-1.2 keV, center: 1.2-2.0 keV, right: 2.0-7.0 keV) according to eq. (3.19).	106
B.7	Synthetic data reconstruction uncertainties for the individual energy bins (left: 0.5-1.2 keV, center: 1.2-2.0 keV, right: 2.0-7.0 keV. Top row: Reconstruction results for the flux in $[s^{-1} cm^{-2}]$ for the individual energy bins. Bottom row: Uncertainty maps for the individual energy bins.	106
B.8	Spatial power spectra of the posterior mean and samples in the diffuse component for each energy bin (red: 0.5-1.2 keV, green: 1.2-2.0 keV, blue: 2.0-7.0 keV)	107

B.9	Posterior means and standard deviations for each energy bin in $[\text{s}^{-1} \text{ cm}^{-2}]$: Top row: Posterior means (red: 0.5-1.2 keV, green:1.2-2.0 keV, blue:2.0-7.0 keV). Bottom row: Posterior standard deviations (left: 0.5-1.2 keV, center:1.2-2.0 keV, right:2.0-7.0 keV).	107
C.1	Visualization of eROSITA data per energy bin from left to right, 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV for TM1 to TM6 from top to bottom. . . .	111
C.2	eROSITA exposure maps summed over all 5 TMs.	112
C.3	Results for single-TM reconstructions for TM1, TM2 and TM3. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively. Continued for TM4 and TM6 in fig. C.4	114
C.4	Results for single-TM reconstructions for TM4 and TM6. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively.	115
C.5	Posterior mean of the sky reconstruction in the different energy bins. . . .	115
C.6	Standard deviation per energy bin for the reconstruction shown in fig. 4.4 and fig. C.5.	115
C.7	Continued in fig. C.8.	116
C.8	Posterior mean of the noise weighted residuals (NWRs) for single-TM reconstructions.	117
C.9	Visualisation of the exposure corrected simulated data per energy bin from left to right, 0.2 - 1.0 keV, 1.0 - 2.0 keV and 2.0 - 4.5 keV for TM1 to TM6 from top to bottom.	118
C.10	Visualization of the simulated sky per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).	119
C.11	Visualization of the reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).	119
C.12	Visualization of the standard deviation of the validation reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV)	119
C.13	Visualization of the standardized error of the validation reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).	119

List of Tables

2.1	Hyper-parameters for correlated field samples shown in fig. 2.2. The reference parameters are denoted by an index r . The other indices correspond to the labels of the sub-figures (a , b , c). Modified hyper-parameter means with respect to the reference field are marked in blue.	27
2.2	Inferred (information field theory (IFT)) and NS (NS stat: statistical, NS det: deterministic) results for the computed Gaussian log-evidence, represented by the mean and the according standard deviation for $n_{\text{live}} \in \{2, 10, 1000\}$ given K sample chains. The ground truth is $\ln Z = -30.87$. The histograms of the distributions of the evidences are shown in fig. 2.4.	32
2.3	Parameters for the spike-and-slab likelihood described in eq. (2.29).	34
2.4	Inferred (IFT) and NS (NS stat: statistical, NS det: deterministic) results for the computed spike-and-slab log-evidence, represented by the mean and the according standard deviation for $n_{\text{live}} \in \{2, 10, 1000\}$ given K sample chains. The ground truth is $\ln Z = -15.47$. The histograms of the distributions of the evidences are shown in fig. 2.5.	35
3.1	Number of hyper-parameters in each model per component	54
3.2	Number of latent parameters in each model per component	54
3.3	Information on the Chandra ACIS observations for the used data of SN1006 according to Winkler et al. (2014). Observations taken by the instrument ACIS-S are followed by the ACIS-I observations.	58
B.1	NWRs (eq. (3.23)) for each dataset in table 3.3 and energy bin.	109
C.1	Flags and their corresponding data types for <code>evtool</code> , where $\text{tmid} \in \{1, 2, 3, 4, 6\}$ and $(\text{emin}, \text{emax}) \in \{(0.2, 1.0), (1.0, 2.0), (2.0, 4.5)\}$	112
C.2	Flags and their corresponding data types for <code>expmap</code> , where $\text{tmid} \in \{1, 2, 3, 4, 6\}$ and $(\text{emin}, \text{emax}) \in \{(0.2, 1.0), (1.0, 2.0), (2.0, 4.5)\}$	112

Chapter 1

Introduction

The human brain is designed to assess situations visually. Since the dawn of humanity, this is how we have processed and visualized the information from our environment, perceived the speed of objects in our field of vision and analyzed the reactions and emotions of other people in our community. Specifically, our eyes process optical light, which is emitted or reflected from our surroundings. Since the 11th century, there has been a great interest in capturing what we see in images, leading to today's digital photography. Although cameras are getting better and better, it is clear that they most likely will never be accurate enough to detect even the tiniest detail of reality. In other words, the resolution of the camera taking a photo will never be able to capture the continuity of our surroundings. In fact, the photograph provides a finite number of data points, defined by the number of pixels, which constrain our assumptions about the imaged object, but do not allow complete knowledge of it. However, in a scientific context, we are usually interested in getting as much information as possible about exactly that continuous ground truth. Therefore, imaging techniques need to be developed that can incorporate additional information to derive the ground truth from the finite data set.

Imaging is in general not just about capturing and analyzing sources of optical light. Imaging methods allow us to create visual representations of any physical quantity of interest. Typically, physical quantities are color coded in such representations, producing images that are easily accessible to human perception. In this way, images allow us to quickly process large and complex data sets. Navigation maps are an everyday example. They translate geographic information into a visual format designed for easy human interpretation. Such navigation maps are usually a two-dimensional imaging example. However, imaging can be used from one-dimensional space, such as visualizing a spectrum, through three-dimensional space, like a color image, to arbitrary dimensions. Hence, imaging techniques, which give us an intuition about patterns and morphology in physical quantities, play an important role in almost all areas of science, from material sciences to astrophysics.

In this work, the main focus is on the imaging of the X-ray sky. Such X-rays originate from high-energy phenomena in the universe, like it is the case in neutron stars, black holes and supernova remnants. These X-rays are measured by telescopes on satellites, because, fortunately for mankind, X-rays do not pass through our atmosphere. In order

to reconstruct the continuous field of X-ray flux from a finite list of data points recorded by the telescope, advanced imaging techniques are required.

Such imaging techniques following a Bayesian approach are discussed in the remainder of this work. In particular, this introductory chapter provides a structured overview of the key concepts and methodologies for imaging continuous fields using Bayes statistics (section 1.1), X-ray astronomy (section 1.2), and the specifics of Bayesian imaging of the X-ray sky (section 1.3). Section 1.4 outlines the objectives and structure of the subsequent chapters and my corresponding five major publications that summarize my work in this field.

1.1 Bayesian imaging of physical scalar fields

Physical *fields* are everywhere in nature. They can be scalar fields, like the temperature field, vector fields, like the magnetic field, or tensor fields, like the metric of general relativity. In general, such a physical field assigns a physical quantity to each point in an n -dimensional space, $z \in \Omega_z \subseteq \mathbb{R}^n$. Imaging is all about creating a visual representation of such physical quantities across the image domain Ω_z . Continuing with the example of temperature fields, imaging generates a visualization of a temperature map. Ideally, one would like to have as complete information as possible about the temperature, $T(z)$, at each point in space, z , so that the image can be displayed in arbitrarily high resolution. However, only a finite set of data points can be measured, which imposes constraints on the underlying temperature field. In fact, the temperature field assigns a temperature to every point in the image domain and thus has an infinite number of degrees of freedom. This leads to an under-constrained problem and an infinite number of possible field configurations that are consistent with the data. Finally, one should try to extract the most plausible of these. This example of the challenge of imaging temperature fields is broadly generalizable to imaging any physical field.

In the following we will assume, as is most common in image processing, that we are interested in a scalar field, $s : z \rightarrow s(z) \in \Omega_s \subseteq \mathbb{R}$, such as the temperature field, for which we can assign a scalar, $s(z)$, to each point z . In case s is our quantity of interest, we call it the signal field. In particular, we want to get the most information about s , under the constraints of a discrete *data* set d . When trying to infer an infinite number of degrees of freedom of s from a limited number of constraints, the resulting inferences can only be expressed probabilistically. Here, information field theory (IFT) (Enßlin, 2013, 2019) combines a probabilistic approach with field theory to infer and visualize the actual field of interest given the data. The theoretical framework for visualizing scalar physical fields using IFT is introduced below.

Imaging using digital computers is usually done on a finite number of pixels. In other words, s is displayed on a particular grid. In particular, in the mathematical framework of IFT the continuum limit from a pixelated signal array to a continuous signal field is taken. This allows the signal to be inferred from the data as a continuous field, independent of the final resolution or pixelization of the used, digital field realization. To update the knowledge

about the signal after receiving the data, IFT leverages Bayesian inference. Here, the state of knowledge is described by probability densities. In particular, each possible field configuration is assigned a value which is the corresponding probability density of the real field being in infinitesimal distance to that configuration. Similarly, the initial belief state about the signal is described by a probability density $\mathcal{P}(s = \bar{s})$, where $\mathcal{P} \in [0, \infty)$, which is the probability density that the unknown field s has the configuration \bar{s} . In fact, the belief state in s usually depends on *contextual information*, I , which could be any knowledge about the underlying system that is not considered as data, such as the assumption that there should be no abrupt changes in temperature over short distances. Accordingly, this contextual information must be included in the probabilistic description. With this in mind, the abbreviated notation $\mathcal{P}(s) := \mathcal{P}(s = \bar{s}|I)$ will be used in the following. In Bayes' inference, $\mathcal{P}(s)$ is called the *prior* or prior probability, describing the knowledge about s before any data is taken into account. For a number of reasons, that are detailed later, it makes sense to refer to $\mathcal{H}(s) = -\ln \mathcal{P}(s)$ as the (prior) information Hamiltonian, information potential, or simply *information* on s .

Bayes' theorem states how to update the prior to a knowledge state which is data informed, called the *posterior*, $\mathcal{P}(s|d) := \mathcal{P}(s = \bar{s}|d, I)$. The posterior, is mathematically described by the probability density that s takes a certain configuration \bar{s} given the contextual information and data. In particular, *Bayes' Theorem*,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(s, d)}{\mathcal{P}(d)} = \frac{\mathcal{P}(s)\mathcal{P}(d|s)}{\mathcal{P}(d)}, \quad (1.1)$$

utilizes the prior and the, so called *likelihood*, $\mathcal{P}(d|s) := \mathcal{P}(d|s = \bar{s}, I)$, in order to calculate the posterior. The likelihood, is the probability that the data was measured given a certain field configuration. Thereby, it models the measurement process. Finally, the posterior needs to be normalized, in order to be a proper probability density function. The resulting, normalization $\mathcal{P}(d) = \int \mathcal{D}s \mathcal{P}(s, d)$ is referred to as the *evidence*. In terms of information functions, Bayes' theorem becomes,

$$\mathcal{H}(s|d) = \mathcal{H}(s) + \mathcal{H}(d|s) - \mathcal{H}(d), \quad (1.2)$$

stating that the above defined mathematical concept of information is an additive, in physical language, extrinsic quantity.

Bayes theorem allows one to reason about s under incomplete or corrupted data by combining the knowledge we already have with the measured data. The result is not only a point estimate on s , but the whole posterior distribution over all possible configurations of s or some approximation of it. Thereby, any quantity of interest which can be written as an posterior expectation value,

$$\langle \cdot \rangle_{(s|d)} = \int \mathcal{D}s \cdot \mathcal{P}(s|d) \quad (1.3)$$

like the posterior mean $\langle s \rangle_{(s|d)}$ or variance $\langle (s - \langle s \rangle_{(s|d)})(s - \langle s \rangle_{(s|d)})^\dagger \rangle_{(s|d)}$ can be calculated.

The following outlines the three basic steps of Bayesian inference, as noted in Gelman et al. (2013): First, section 1.1.1 describes the probability models, including the modelling

of the prior and the likelihood. Second, section 1.1.2 describes challenges in inference and possible approximations of the posterior conditioning on the data. Third, and finally, ways to validate an appropriate approximation of the posterior are discussed in section 1.1.3. All the models described are implemented using the software packages Numerical Information Field TheorY (NIFTy) (Arras et al., 2019; Selig et al., 2013) and its further JAX-based reimplementations (Edenhofer et al., 2024). NIFTy discretizes the continuous fields for computational use.

1.1.1 Probability models

Theoretical models give us a simplified representation of the underlying reality - as do probability models of our current state of knowledge. The signal itself and its measurement process can be arbitrarily complex. In the probability models of prior and likelihood, the goal is to describe the state of belief about the signal and the measurement, including the most important patterns. In the end, the prior and the likelihood model determine the unnormalized posterior, i.e. $\mathcal{P}(s|d) \propto \mathcal{P}(s, d) = \mathcal{P}(s)\mathcal{P}(d|s)$. How to derive an approximation to the true posterior or compute the evidence as the normalization is described in section 1.1.2.

The prior model for s is supposed to describe the assumptions about the signal before looking at any data. In Bayesian inference, the prior model is essential to reason about the continuous signal under the uncertainty introduced by the limited constraints of finite data. However, inaccurate prior assumptions can hinder inference and lead to incorrect posterior estimates of s . Therefore, the prior model must be carefully chosen and well justified.

One approach would be to not a priori prefer any possible field configuration for s , or in a mathematical sense to choose $\mathcal{P}(s)$ to be flat. This is called a *non-informative prior*, as the prior does not contribute any information about the signal during inference. In fact, in this case the posterior knowledge about the signal is inferred entirely from the data. However, as noted in Gelman et al. (2013), a flat prior is not flat in every coordinate system. This can be illustrated on the example of the temperature map of an area of the Earth. Suppose the prior for the temperature field, $T(\mathbf{x})$, does not favor any configuration for positive temperatures, i.e. $\mathcal{P}(T) \propto c = \text{constant}$ for $T \in \mathbb{R}^+$ is assumed. In statistical thermodynamics, one instead considers the reciprocal temperature, $\beta = \frac{1}{T}$. Given that the prior normalization must remain the same,

$$1 = \int \mathcal{D}T \mathcal{P}(T) = \int \mathcal{D}\beta \mathcal{P}(\beta), \quad (1.4)$$

$$\rightarrow \lim_{n \rightarrow \infty} \prod_i^n \int dT(\mathbf{x}_i) c = \lim_{n \rightarrow \infty} \prod_i^n \int d\beta(\mathbf{x}_i) \mathcal{P}(\beta(\mathbf{x}_i)), \quad (1.5)$$

$$= \lim_{n \rightarrow \infty} \prod_i^n \int dT(\mathbf{x}_i) \beta(x_i)^2 \mathcal{P}(\beta(x_i)), \quad (1.6)$$

the prior for the reciprocal temperature field, is no longer flat but informative, $\mathcal{P}(\beta) \propto \frac{1}{\beta^2}$. As a result, Bayesian inference must always involve a specific formulation of prior knowledge. For example, Jeffrey (Jeffreys, 1946) worked out an invariant form of the prior probability distribution defined by the square root of the Fisher information.

In the following prior models are considered, which allow to constrain the degrees of freedom of the inference problem, which are not set by the data. However, these constraints shall be set to sensible bounds, while still being indifferent about solutions which are a priori equally plausible. The fact that the prior densities depend on the coordinate system chosen, as shown above, can be used to our advantage. Any prior, no matter how complex its structure, from high-dimensional to multi-modal, can be described by a simpler probability distribution accompanied by a coordinate transformation. In this case, the coordinate transformation maps between the space on which the simpler probability distribution is defined and the coordinate space of the more complex one.

Following this principle, the prior models used in this paper are based on standard Gaussian priors and a corresponding mapping f from the standard normally distributed field $\xi \sim \mathcal{N}(\xi|0, 1)$ to the signal $s = f(\xi)$ with a more complex prior model. The corresponding technique was introduced in Knollmüller and Enßlin (2020) and is based on *inverse transform sampling* (Devroye, 1986). Accordingly, and as described in Enßlin (2022), the prior model is implemented as a *generative model*. The a priori standard Gaussian distributed field ξ , is called the *latent field*. From now on, we are interested in obtaining an estimate of the latent field and its parameters. The application of the mapping $f(\xi)$ yields the corresponding estimate of the signal s from the hidden latent field.

The likelihood model aims to capture the essential parts of the measurement process. In analogy to the generative model introduced for the prior, this should describe the likelihood of the data given s as a function of the hidden latent field ξ . As noted in Gelman et al. (2013) this leads to Bayesian *hierarchical models* with the joint probability density,

$$\mathcal{P}(s, \xi, d) = \mathcal{P}(\xi) \mathcal{P}(s|\xi) \mathcal{P}(d|s, \xi). \quad (1.7)$$

Marginalizing over s gives the joint probability of the latent field and the data via the hyper-prior, $\mathcal{P}(\xi)$, and a hierarchical likelihood model, $\mathcal{P}(d|f(\xi))$,

$$\mathcal{P}(\xi, d) = \int \mathcal{D}s \mathcal{N}(\xi|0, 1) \delta(s - f(\xi)) \mathcal{P}(d|s, \xi) = \mathcal{N}(\xi|0, 1) \mathcal{P}(d|f(\xi)). \quad (1.8)$$

This allows inferring ξ , which completely encodes the signal field via the transformation f . The likelihood model is defined hierarchically by $\mathcal{P}(d|s) = \mathcal{P}(d|f(\xi))$ and describes, as the notation suggests, the likelihood that d was measured given the a priori standard normal latent ξ field describing s .

The measurement model may be described by a function, \mathcal{M} , capturing all the effects of the production of data from the signal. Usually, two factors are taken into account: (i) The instruments response, \mathcal{R} , which is an simplified model how the instrument under consideration would transform the signal into data. Therefore, it must map from signal

space, Ω_s , to data space. (ii) Nuisance parameters, η , such as noise, which are unpredictable random fluctuations in the measurement. The measurement equation can generally be written down as, $d = \mathcal{M}(\mathcal{R}(s), \eta)$, where \mathcal{M} is an arbitrary function of $\mathcal{R}(s)$ and η . The joint probability of the data and the nuisance parameters given the signal is then defined via,

$$\mathcal{P}(d, \eta | f(\xi)) = \mathcal{P}(d | f(\xi), \eta) \mathcal{P}(\eta | f(\xi)) = \delta(d - \mathcal{M}(\mathcal{R}(f(\xi)), \eta)) \mathcal{P}(\eta | f(\xi)). \quad (1.9)$$

The hierarchical likelihood model, which includes the discretization of the signal as well as all the information given by the data, can be derived from eq. (1.9) via marginalization over η ,

$$\mathcal{P}(d | f(\xi)) = \int d\eta \delta(d - \mathcal{M}(\mathcal{R}(f(\xi)), \eta)) \mathcal{P}(\eta | f(\xi)). \quad (1.10)$$

1.1.2 Posterior approximation

Given probabilistic models for the prior and the likelihood, the product rule allows to construct the unnormalized posterior,

$$\mathcal{P}(s | d) \propto \mathcal{P}(s, d) = \mathcal{P}(d | s) \mathcal{P}(s). \quad (1.11)$$

However, to obtain a proper posterior, its normalization is required. This is defined by what is known as Bayesian evidence,

$$\mathcal{P}(d) = \int \mathcal{D}s \mathcal{P}(s, d) = \int \mathcal{D}s \mathcal{P}(s) \mathcal{P}(d | s). \quad (1.12)$$

If the prior and the likelihood are conjugate, which means that their combination results into a posterior of the same family of probability distributions as the prior, a closed-form solution for the evidence exists. In other words, an analytical solution for the posterior is available. A *family* of probability distributions, \mathcal{P}_ϕ , are probability distributions that share the same functional form, but differ in the value of their parameters ϕ , like Gaussian distributions that differ in their mean and variance. A simple example here is a Gaussian prior, $s \sim \mathcal{N}(s | 0, S)$, combined with a linear measurement equation, $d = Rs + n$, with Gaussian noise, $n \sim \mathcal{N}(n | 0, N)$ and a linear response $\mathcal{R}(s) = Rs$. For this scenario the unnormalized posterior,

$$\mathcal{P}(s, d) = \mathcal{N}(s | 0, S) \int dn \delta(d - Rs - n) \mathcal{N}(n | 0, N) = \mathcal{N}(s | 0, S) \mathcal{N}(d - Rs | 0, N), \quad (1.13)$$

turns out to be a Gaussian as well.

However, non-conjugate models are often needed in order to model more complex relations between the prior and the posterior, leading to an intractable normalization of the posterior. In this case, the true posterior is inaccessible and must be approximated in some way. In the literature, there are two main approaches discussed in order to approximate the posterior: Approximating the posterior by a set of posterior samples or finding an

analytic distribution, that is as close as possible to the true posterior. A brief overview of these methods is given in chapter 2.

In this work, the focus is on approximating the posterior, $\mathcal{P}(\xi|d)$, with a simpler family of distributions, $\mathcal{Q}_\phi(\xi)$, where ϕ denotes its variational parameters. A common example of a family of approximate distributions is the Gaussian distribution. Once one has chosen a family of distributions, one needs a measure of the distance between the actual posterior $\mathcal{P}(\xi|d)$ and its approximation $\mathcal{Q}_\phi(\xi)$. This distance measure has to be minimized, in order to find the optimal parameters, ϕ . In variational inference (VI) (Blei et al., 2017) the Kullback–Leibler divergence (KL) (Kullback and Leibler, 1951),

$$D_{\text{KL}}(\mathcal{Q}_\phi(\xi)||\mathcal{P}(\xi|d)) = \int d\xi \mathcal{Q}_\phi(\xi) \ln\left(\frac{\mathcal{Q}_\phi(\xi)}{\mathcal{P}(\xi|d)}\right) \quad (1.14)$$

$$= \langle \mathcal{H}(\xi, d) \rangle_\phi - \langle \mathcal{H}_\phi(\xi) \rangle_\phi - \mathcal{H}(d), \quad (1.15)$$

is used as the distance measure, where $\mathcal{H}_\phi = -\ln \mathcal{Q}_\phi$ is the information content of \mathcal{Q}_ϕ and $\langle \cdot \rangle_\phi$ is the average with respect to \mathcal{Q}_ϕ . Intuitively, the KL gives a measure of how much information is invented on average when the posterior is approximated by a given $\mathcal{Q}_\phi(\xi)$. While minimizing the KL in eq. (1.15) with respect to the variational parameters ϕ , the log-evidence term, $\mathcal{H}(d)$, can be neglected, since it is an additive constant in ϕ . As a result, VI allows one to approximate the posterior without having to compute the evidence.

There are several approaches to VI that differ, among other things, in the parametric family of probability distributions chosen for the approximation. A common approach is to approximate the posterior by a Gaussian, called Gaussian VI. In this case, only two variational parameters need to be inferred, namely the mean and the variance. In particular, metric Gaussian variational inference (MGVI) (Knollmüller and Enßlin, 2020) builds the Gaussian VI on top of generative prior models, as discussed above. In this study, geometric variational inference (geoVI) (Frank et al., 2021) is used which, instead of solely learning the variational parameters of a given parametric probability family, minimizes the KL for a coordinate transformation to a space in which the posterior is Gaussian distributed. This generates a problem adapted parametrization. In this way, more complex distributions, independent of a static parametric family, can be approximated.

In certain instances, the focus is explicitly on computing the evidence. For example, when comparing models. Model comparison distinguishes between multiple models representing hypotheses about the signal structure. By analogy with the temperature field example, two nearby points in the temperature maps could be assumed to be completely independent in one model and correlated in another. Model comparison is a tool for the identification of the most appropriate model among these options within the data constraints. The desired quantity is the probability for the model given the data,

$$\mathcal{P}(M|d) = \frac{\mathcal{P}(M)\mathcal{P}(d|M)}{\mathcal{P}(d)}, \quad (1.16)$$

where $\mathcal{P}(d) = \sum_M \mathcal{P}(d|M)\mathcal{P}(M)$ is the *meta-evidence*, i.e. the probability for the data regardless of the model used. This probability distribution may be used to compare the

compatibility between two models M_1 and M_2 using the *Bayes factor*,

$$B_{12} = \frac{\mathcal{P}(M_1|d)}{\mathcal{P}(M_2|d)} = \frac{\mathcal{P}(d|M_1)\mathcal{P}(M_1)}{\mathcal{P}(d|M_2)\mathcal{P}(M_2)}. \quad (1.17)$$

Under the assumption of equal prior probability distributions for each of the models, i.e. $\mathcal{P}(M_1) = \mathcal{P}(M_2)$, the comparison of models is determined solely by their evidences, $\mathcal{P}(d|M_1)$ and $\mathcal{P}(d|M_2)$. A scientific application of a model comparison is described in Handley (2021), where the evidence calculated from Planck data is used to distinguish between different cosmological models.

Therefore, when comparing models, it is necessary to calculate the integral in eq. (1.12). In chapter 2 we describe several approaches to do this, focusing on one of them called nested sampling (NS). NS is an algorithm that transforms the high-dimensional integral in eq. (1.12) into a one-dimensional integral over the contours of the likelihood and the corresponding enclosed prior volume. This procedure introduces a stochastic error into the calculation of the evidence, as will be discussed later.

1.1.3 Model evaluation

As mentioned above, the key components for field inference are the prior and likelihood models. However, these need to be chosen carefully. Incorrect descriptions of either can lead to significant problems in the inference and in the final estimates and conclusions. Fortunately, two key aspects of the inference presented here offer numerous ways to analyze the reconstruction: the statistical approach and the use of generative prior models. Due to the statistical approach, the final result is a whole set of posterior samples, which allows the calculation of any posterior sample average of interest, such as the uncertainty in terms of standard deviation. These measures are an important analytical tool and will be used to interpret the validity of the reconstruction in the subsequent chapters.

To check the overall consistency of the implementations of the likelihood models and the inference scheme, all models are tested on simulated data. The generative nature of the prior model allows us to draw prior samples, pass them through the instrument description, and incorporate possible noise, allowing us to generate simulated data. On the basis of these synthetic data, we do Bayesian inference and compare the results to the ground truth signal using several diagnostic measures, such as various residuals, to ensure consistency.

1.2 X-ray astronomy and its observations

Astronomy is the study of celestial objects and phenomena. It has been studied by humans for thousands of years, from observing the sky with the naked eye to using telescopes beyond the Earth's atmosphere. High-energy astronomy is an engaging branch of astronomy that studies the most energetic processes in the universe. Such processes radiate high-energy radiation, such as X-rays or γ -rays. While our atmosphere is almost transparent in the energy range of optical light, X-rays and γ -rays are absorbed by the atmosphere and are

therefore not measurable at the Earth's surface (Seward and Charles, 2010). Accordingly, before the development of space technology, there was no awareness of the high energy phenomena in the universe.

X-ray astronomy is the study of all objects and processes that emit X-rays, i.e. radiation in the range 0.1-100 keV. Section 1.2.1 summarizes the main underlying processes for X-rays from astronomical objects. The energy range of X-rays not only makes them impossible to be observed from Earth, but also complicates the optics needed to reflect and focus them on a detector. How state-of-the-art telescopes meet this challenge is discussed in section 1.2.2.

1.2.1 X-ray emitting astronomical objects and there emission processes

As noted in Seward and Charles (2010), there are essentially three predominant emission processes that produce X-rays: Bremsstrahlung, Synchrotron radiation and blackbody radiation. These emissions differ not only in their underlying physical processes, but also in their spectral shape. Thus, by measuring the spectrum of a celestial source, it is possible to derive the processes involved and other physical conditions, such as temperature and elemental composition, in the observed region.

The different types of emissions are explained below, with examples of their occurrence in supernova remnants (SNRs). A luminous explosion or collapse at the end of the life cycle of a star is called a supernova (SN). As the name suggests, a SNR is the structure left behind after the SN event. Chapter 3 focuses on the imaging of a particular SNR called SN1006. SN1006 is expected to be the remnant of a Type Ia SN. When a small star has exhausted its nuclear fuel and becomes a white dwarf, this type of supernova can occur. Under these circumstance the Pauli principle prevents the white dwarf from further collapsing, called the electron degeneracy pressure. A commonly accepted model assumes that this white dwarf is in a binary system with a companion star. In this setting, the white dwarf is assumed to accrete mass from the companion star, eventually reaching a mass such high that gravity overcomes the resistance of the electron density pressure. Typically, the result is a collapse of the white dwarf to a neutron star, known as a Type Ia SN.

Bremsstrahlung is well known from the discovery of X-ray emission by Röntgen in 1895. It generally occurs when charged particles are accelerated in the electromagnetic field of other charged particles. As a result, the accelerated particle releases energy in the form of high energy photons, called bremsstrahlung. In astronomy, bremsstrahlung occurs in a hot gas with temperatures $T > 10^5 K$. Under these circumstances, some of the atoms of the hot, thin gas are ionized (Seward and Charles, 2010) and electrons are accelerated by the ionized atoms. They emit X-rays.

At temperatures where not all atoms are ionized, additional characteristic line emission occurs. Some of the electrons will transfer energy to the not ionized atoms, resulting in an

excited state of the atom. As soon as the electrons of the atom return to their ground state, energy in form of a photon characteristic to the energy levels of the atom is released, called characteristic line emission. The resulting spectrum composed out of bremsstrahlung and characteristic line emission depends on, among other things, the temperature of the hot gas and its atomic composition. Consequently, its measurement provides insight into the physical conditions in the observed region.

During the SN stellar material is ejected into space. The shell of the ejected material sweeps up the the surrounding medium of the former white dwarf leading to a shock front. As the shock front expands, the medium surrounding the SN heats up and ionizes. Bremsstrahlung and characteristic line emission occur, which are often referred to as the *thermal emission* of the SNR.

Synchrotron radiation results from the acceleration of relativistic electrons in a magnetic field. As relativistic electrons are deflected by a magnetic field, they emit radiation with a broad energy range. The resulting spectrum of the synchrotron radiation is characterized by a power law defined via the exponent, called the spectral index α .

In certain cases, synchrotron emitting regions are located along the shock fronts of SNRs, usually called the *non-thermal emission* of the SNR. It was this observation that provided the first clue to the origin of cosmic rays. For a long time, the question of where these energetic particles come from and where they are accelerated remained unsolved. The fact that electrons must have an energy of 100 TeV to produce X-ray synchrotron emission (Seward and Charles, 2010) may indicate that not only electrons but also positive ions could be accelerated to such high velocities at the shock front. Eventually, cosmic rays could be produced.

Blackbody radiation is, as the name suggests, emitted by a black body. A black body is an object that absorbs any electromagnetic radiation and then emits it with a spectrum characterized by the temperature of the black body - the higher the temperature, the higher the energy of the blackbody radiation. The final spectrum of the blackbody radiation is a continuous spectrum, which is mathematically described by Planck's law of blackbody radiation (Rybicki and Lightman, 1991). An example of celestial objects, emitting blackbody radiation are neutron stars, which have extremely high surface temperatures.

1.2.2 X-ray telescopes

X-ray telescopes have special requirements in two ways. First, the telescopes must be operated in space due to the absorption of X-rays in the Earth's atmosphere. Accordingly, the size and weight of the instruments are limited. Second, X-ray optics in telescopes have specific needs due to the interaction of X-rays with matter. In particular, X-rays are primarily absorbed by matter rather than reflected. As noted in Arnaud et al. (2011), both of those requirements lead to the need for a compromise between the size of the angular

area that the telescope can observe, i.e. the field of view (FOV), and a high angular resolution.

Even if X-rays generally penetrate a material, they may be reflected at small grazing angles. At very small grazing angles, the incoming X-rays may even be completely reflected. This is called *total external reflection* and the corresponding grazing angle beyond which total external reflection occurs is denoted as the *critical angle* (Arnaud et al., 2011). If the critical angle is small, the reflecting mirrors can only collect X-rays from a small region of the sky. Therefore, it is desirable to have a critical angle as large as possible, which is proportional to the atomic number of the mirror's material.

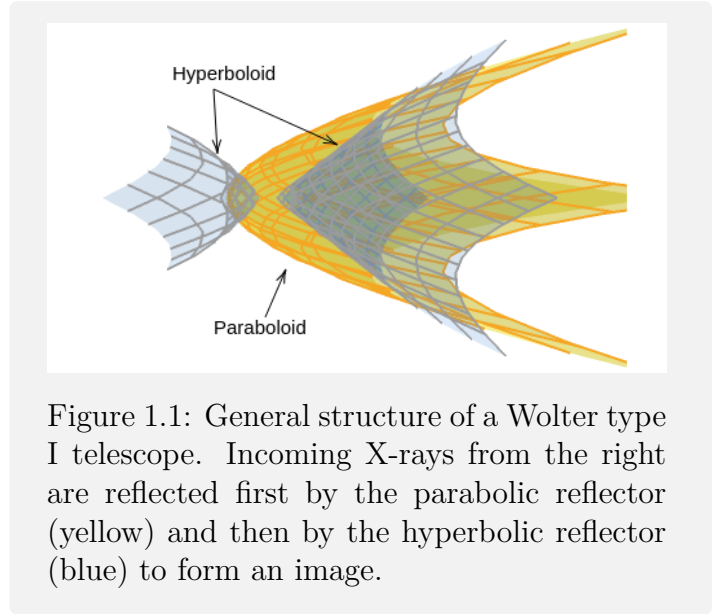


Figure 1.1: General structure of a Wolter type I telescope. Incoming X-rays from the right are reflected first by the parabolic reflector (yellow) and then by the hyperbolic reflector (blue) to form an image.

In Wolter (1952) it was first noted that two reflections are required to focus X-rays with a finite FOV. In particular, three possible geometries, Wolter type I-III, were introduced for focusing X-rays onto the detector using different configurations of a hyperbolic and a parabolic reflecting surface. Most X-ray telescopes, including eROSITA, Chandra and XMM-Newton, use the Wolter type I geometry. The general structure of the hyperbolic and parabolic reflecting surfaces in a Wolter type I mirror is shown in fig. 1.1. In this case, the incoming X-rays from the right side of the image are first reflected at the inner parabolic surface, then at the inner hyperbolic surface, and thereby focused on the detector.

In addition, a diffraction grating (Heilmann et al., 2024) is often placed between the mirror assembly and the detector to split the converging X-ray beam into its spectral components. Subsequent X-ray detectors, particularly charge-coupled devices (CCDs), measure individual X-ray photons as events by converting each incoming photon into an electrical charge by *photoelectric absorption*.

In the following, the specific X-ray telescopes Chandra and eROSITA are considered. While Chandra has the highest angular resolution achieved so far for X-ray telescopes, and focuses on the observation of objects with small extents, eROSITA with its large FOV focuses on the observation of large structures and all-sky surveys. The specifics of these individual instruments are described in more detail in the corresponding chapters, chapter 3 and chapter 4.

1.3 Bayesian X-ray imaging

Once data is received from one of the telescopes, the goal is to extract as much information as possible from the potentially noisy, corrupted data and update our knowledge of the observed object of interest. However, there are a number of challenges in this process. In section 1.2.1 it was discussed that there are multiple emission processes for X-ray emission. Likewise, there are usually multiple sources with different emission processes in an observation. Sometimes all of these sources are the signal of interest, while at other times we may want to construct only one of them separately. Any of the components that are of interest are referred to as *signal components*, and all sources that are not of interest as the *background*. Adding to the difficulty, each of these components may have non-trivial spatial and spectral correlation structures about which we have little knowledge.

In addition, the measurement process described in section 1.2.2 needs to be understood for each telescope in order to deduce how to build the likelihood model. Usually, in X-ray imaging, these include the measurement noise, the telescope point spread function (PSF) and the exposure. Imperfections in the optics lead to a spread out of point sources which is described mathematically by the so-called PSF. The *exposure* is commonly defined as the product of the observation time and the telescope's collection capacity, called the *effective area*. The conceptual aim is to be able to invert the measurement process. In other words efficient deconvolution, exposure correction, and noise reduction techniques need to be developed. Particular difficulties arise when multiple observations of overlapping regions of the sky are to be combined. For example, a point source observed once may have different measurement effects when observed a second time from a slightly different angle. This is especially the case if the PSF varies significantly over the FOV, denoted as *spatial variant PSF*.

These challenges can be tackled using Bayesian imaging for fields, namely IFT. We consider the X-ray sky as the continuous signal field, s , which may have several components, like diffuse emission and point sources. By building individual and different prior models for the signal components, we can infer them jointly from the data and finally decompose them. These components may be assumed to have a priori different spatial and spectral correlation structures. These correlation structures are a priori unknown, but one can make several a priori assumptions on them. Given these prior assumptions the correlation structure is reconstructed concurrently with the signal. Two implementations of such prior models for the X-ray sky are discussed in chapters 3 and 4.

The CCDs of the X-ray telescope measure the incoming X-ray flux from the signal as photon counts. The number of counts in each pixel of the CCD is affected by the PSF and the exposure. Thus, among others, the PSF and the exposure should be included in the instrument response, \mathcal{R} . Since the measured photon counts describe discrete events, the photon counts are Poisson distributed, often referred to as Poisson or shot noise. Therefore, the likelihood for X-ray imaging is modelled by a Poisson distribution for the data d given

the signal s ,

$$\mathcal{H}(d|\mathcal{R}(s)) = \sum_{i=1}^n (\mathcal{R}(s))^i - d^i \ln(\mathcal{R}(s))^i + \ln(d^i!), \quad (1.18)$$

where n is the number pixels in the data. Besides, the probabilistic description allows to take into account a set of m observations. Under the assumption of independent measurements, the likelihoods for each observation factorize. Accordingly, the full likelihood information is defined via the sum of the individual likelihoods in eq. (1.18),

$$\mathcal{H}(\{d_j\}|\mathcal{R}(s)) = \sum_{j=1}^m \sum_{i=1}^n (\mathcal{R}(s))_j^i - d_j^i \ln(\mathcal{R}(s))_j^i + \ln(d_j^i!). \quad (1.19)$$

Eventually, using one of the inference schemes proposed in section 1.1.2, we obtain a set of posterior samples that allow to compute among others the mean of the signal component and its uncertainty.

1.4 Objectives and structure of the thesis

The main objective of the presented work is to bridge the gap between methodological innovation and application by leveraging Bayesian methods for model comparison and imaging in X-ray astronomy. My five major contributions consist of three refereed first-author publications - two articles (Westerkamp et al., 2024a,b) and one conference proceedings paper (Westerkamp et al., 2023) - and two joint first-author publications under review (Eberle et al., 2024b,a). Each of the works features a recurring challenge: The ill-posed problem of reconstructing a continuous field from a discrete data set.

The first part of the thesis (chapter 2) focuses on methodological advances for model comparison. In particular, an algorithm and its implementation are presented to improve the calculation of nested sampling evidence using Bayesian field inference of a one-dimensional field. The initial idea of the algorithm is published in the conference proceedings Westerkamp et al. (2023).

The subsequent chapters utilize Bayesian field inference for imaging of fields in two- to three-dimensional space from X-ray data. The ultimate goal of the project is the spatio-spectral large-field reconstruction of the X-ray sky in an accelerated framework.

The first spatio-spectral Bayesian X-ray imaging approach is shown in chapter 3. Here, a spatio-spectral prior model in analogy to Scheel-Platz et al. (2023) is implemented for X-ray imaging the first time and applied to the Chandra data of SN1006. A corresponding likelihood modelling the features of Chandra in terms of the exposure and a spatially invariant PSF is built. This likelihood model allows to incorporate the data taken from in total eleven observations of the remnant. A multi-step approach is introduced where the spatial reconstruction obtained for a single energy bin is used to derive an informed starting point for the full spatio-spectral reconstruction in order to speed up the imaging process. The result is the first denoised, decomposed and deconvolved image of SN1006 inferred

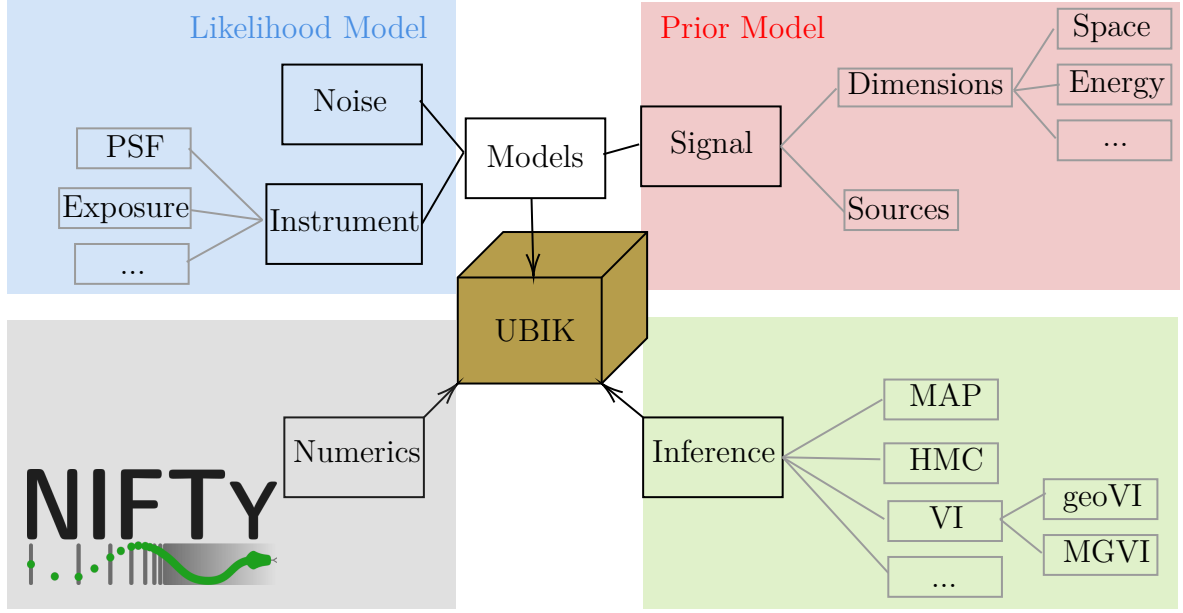


Figure 1.2: Core components of UBIK- The Universal Bayesian Imaging Kit.

via Bayesian methods from Chandra data. This section is published in Westerkamp et al. (2024a).

In chapter 4 the results of collaborative work on X-ray imaging for eROSITA with Vincent Eberle, Matteo Guardiani and myself as first authors are presented. In particular, the focus is on spatio-spectral imaging of eROSITA Early Data Release (EDR) data of the Large Magellanic Cloud (LMC) region of SN1987A. In addition to the description of the eROSITA instrument, including the exposure and in this case a spatially varying PSF for each data set, a novel JAX-compatible spatio-spectral prior model is implemented.

The whole project has been focused on the goal of ultimately building a Universal Bayesian Imaging Kit (UBIK) to be applied to X-ray data. In general, UBIK is intended to be a flexible and modular framework to facilitate Bayesian imaging. Chapter 5 presents the UBIK framework in the form of a software release of the software package J-UBIK, the JAX-accelerated Universal Bayesian Imaging Kit. It is based on NIFTy.re, the JAX-accelerated version of NIFTy. Figure 1.2 highlights the key components of the Bayesian imaging process that utilize IFT and are incorporated into UBIK. While, the numerics and inference algorithms are already included in the NIFTy software, UBIK focuses on facilitating the construction of likelihood and prior models. The corresponding package is open source and includes both, tools to facilitate the construction of new prior and likelihood models as well as already constructed likelihood and prior models for various types of celestial objects and instruments. At the moment, the models for Chandra, described in chapter 3, and for eROSITA, shown in chapter 4, have been implemented.

1.5 Additional work

In addition to my primary research, I completed a project based on my Master's thesis during the first year of my PhD, which focused on the theoretical analysis of dynamical field inference. Dynamical field inference focuses on the challenge of reconstructing a field that evolves dynamically driven by a stochastic force and is accordingly described by a stochastic differential equation (SDE). As soon as the SDE becomes non-linear in the field, the corresponding reconstruction becomes non-trivial due to terms involving a field-dependent functional determinant. This paper deals with finding an appropriate path integral representation of these problematic terms. It was first published in Westerkamp et al. (2021) as a result of my master's thesis and further developed in the conference proceedings Westerkamp et al. (2022).

Chapter 2

Field based Bayesian evidence inference from nested sampling data

*The following chapter was published in **Entropy** with me as first author (Westerkamp et al., 2024b). A first idea has already been published as a conference paper with me as first author (Westerkamp et al., 2023). I contributed through the theoretical calculations for the Bayesian models, the development of the open source package ¹ and the writing of the manuscript. The project is a result from many discussions and close collaboration with Jakob Roth. In particular, Jakob Roth and I jointly developed and implemented the first prototype of the inference pipeline. This work would not have been possible without many valuable discussions with and constructive feedback from Torsten Enßlin and Philipp Frank on the research direction and interpretation of the results. The scientific exchange with Will Handley for deeper insights into nested sampling was very beneficial to the present work. All authors read, commented, and approved the final manuscript.*

For consistency within this thesis, some of the parameters have been renamed and some of the figures have been adapted according to the layout.

¹The corresponding implementation is public and available at <https://gitlab.mpcdf.mpg.de/ift/public/iftns> (accessed November 28, 2024).

Nested sampling (NS) is a stochastic method for computing the log-evidence of a Bayesian problem. It relies on stochastic estimates of prior volumes enclosed by likelihood contours, which limits the accuracy of the log-evidence calculation. We propose to transform the prior volume estimation into a Bayesian inference problem, which allows us to incorporate a smoothness assumption for likelihood–prior–volume relations. As a result, we aim to increase the accuracy of the volume estimates and thus improve the overall log-evidence calculation using NS. The method presented works as a post-processing step for NS and provides posterior samples of the likelihood–prior–volume relation, from which the log-evidence can be calculated. We demonstrate an implementation of the algorithm and compare its results with plain NS on two synthetic datasets for which the underlying evidence is known. We find a significant improvement in accuracy for runs with less than one hundred active samples in NS but a proneness for numerical problems beyond this point.

2.1 Introduction

In Bayesian inference, we update our knowledge about a measure of interest, which we call the signal, s , on the basis of some given data, d . This signal, s , can be, among other things, a set of model parameters θ_M , the model itself M , or a continuous field φ . Given the prior $\mathcal{P}(s)$, which describes our a priori knowledge about $s \in \{\theta_M, M, \varphi, \dots\}$, the likelihood $\mathcal{P}(d|s)$ of the data given the signal and the normalizing constant $\mathcal{P}(d) = \int ds \mathcal{P}(d|s)\mathcal{P}(s)$, known as the evidence, Bayes' theorem returns the posterior of the signal given the data,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\mathcal{P}(s)}{\mathcal{P}(d)}. \quad (2.1)$$

In particular, Bayesian parameter estimation is concerned with updating knowledge about a set of model parameters given the data. Here, the model parameters in general depend on the given model. Accordingly, the quantity of interest is the posterior $\mathcal{P}(\theta_M|d, M)$. In the case of field inference, we aim to reconstruct a continuous field from a finite dataset by approximating the posterior probability $\mathcal{P}(\varphi|d)$. In fact, in both cases, Bayesian parameter estimation and Bayesian field inference, it is in general not necessary to compute the evidence, as long as one is interested in computing posterior expectation values. In contrast, in model comparison the evidence is the main measure of interest. In model comparison, multiple models, each with parameters and assumptions, are compared by the probability of a model M_i given the data,

$$\mathcal{P}(M_i|d) = \frac{\mathcal{P}(d|M_i)\mathcal{P}(M_i)}{\mathcal{P}(d)}, \quad (2.2)$$

with $\mathcal{P}(d) = \sum_j \mathcal{P}(d|M_j)\mathcal{P}(M_j)$. Assuming that all models have the same a priori probability, this leads to a comparison of the evidences $\mathcal{P}(d|M_j)$ between a set of different

models $\{M_j\}$. The ratio of the evidences of two models is called the Bayes factor, giving the betting odds for or against one of the models compared to the other.

A number of algorithms for Bayesian parameter estimation already exist, which can be divided into two classes of posterior estimation approaches. One approach approximates the posterior, i.e., it tries to find an analytic distribution that is close to the true posterior. The main approach here is variational inference (VI), which minimizes a distance measure between the analytic distribution and the posterior distribution, for example the Kullback–Leibler divergence (KL) (Kullback and Leibler, 1951), and is often used in field inference. The other approach aims to generate a set of samples of the posterior distribution. This set of samples can be used to approximate the true posterior. The most popular posterior sampling algorithm is Markov Chain Monte Carlo (MCMC). A summary of the basics of MCMC and different implementations is given in Hogg and Foreman-Mackey (2018). MCMC methods draw samples directly from the posterior, given a likelihood and a prior model. The simplest algorithm for MCMC is the Metropolis–Hastings algorithm (Robert, 2015), which gives a biased random walk through the parameter space depending on a proposal function and the initialization of the algorithm. Still, there are several challenges, such as the tuning of the proposal function and the initialization, which lead to advanced MCMC methods such as ensemble sampling (Lu and Roy, 2017), Gibbs sampling (Geman and Geman, 1984) and Hamiltonian Monte Carlo (Betancourt, 2017).

For model comparison (eq. (2.2)), we are instead interested in the integration of the likelihood over the prior, or in other words in the evidence calculation. Computing Bayesian evidence is challenging in many applications, as discussed in Buchner (2023). The biggest challenges are high-dimensional posteriors with multiple, well-separated modes or plateaus and a high information gain from the prior to the posterior, increasing the amount of time the algorithm spends in the low posterior mass regime. A concise overview of the different approaches for integration is given in Preuss et al. (2007). Two integration methods that smoothly contract the parameter space from the prior to the posterior are simulated annealing (Wegener, 2005) and nested sampling (NS) (Skilling, 2004, 2006). While simulated annealing uses fractional powers of the likelihood to obtain from the prior to the posterior, NS instead takes samples from slices of the posterior and recombines them at the end. In particular, NS transforms the multidimensional problem of integrating the likelihood over the prior into a series of nested volumes defined by likelihood contours and the enclosed prior. In doing so, NS is able to estimate the log-evidence and the posterior samples simultaneously. The NS algorithm is analyzed further from a physical perspective in Habeck (2015).

As summarized in Buchner and Boorman (2024) there are several challenges in NS. First, the computational cost of NS depends on the choice of the prior, i.e. the broader the prior, the higher the computational cost. Second, sampling from the likelihood restricted prior is not trivial. Finally, the rate at which the posterior is integrated is a stochastic quantity for which there is only a probabilistic description. Accordingly, many improvements to NS have been proposed to address some of these challenges. A number of studies have focused on improving the calculation of evidence for likelihoods with peculiar shapes, such as likelihood plateaus. Likelihood plateaus violate the assumption of uniformity in

sampling and lead to ambiguity in the ranking of samples, which often leads to an underestimation of the prior volume contraction and thus to an overestimation of the evidence. Accordingly, Schittenhelm and Wacker (2021) proposed a preprocessing step to correctly handle the plateaus in the likelihood. The main idea is to decompose the parameter sets into disjoint subsets that divide into plateaus and parts on which the usual NS can be performed.

Other studies have aimed to work with a variable rate at which the posterior is integrated, known as dynamic NS (Higson et al., 2018; Speagle, 2020). Dynamic NS is particularly useful for parameter estimation, as standard NS spends a great deal of computational effort navigating to the posterior peak. Ultimately, dynamic NS allows more samples to be placed in regions where we want higher resolution, and less in less interesting regions. However, most of the progress has been focused on improving the sampling process for the likelihood restricted prior sampling (LRPS). There are generally two different approaches that focus on LRPS - rejection sampling, used by MultiNest (Feroz et al., 2009), and chain-based sampling using Markov chains, implemented for example in PolyChord (Handley et al., 2015). In addition, Salomone et al. (2023) noted that standard NS assumes independent prior samples given the likelihood constraint. However, this is usually not the case, leading to a bias in the evidence calculation. Accordingly, they introduce nested sampling via sequential Monte Carlo (NS-SMC) (Moral et al., 2006) based on the idea of importance sampling, which does not require the imprecise assumption of independent samples.

The dominant error in the evidence calculation, which is based on the statistical estimate of the shrinkage ratio, can be reduced by taking a larger number of samples, which is where improvements in LRPS focus. We take an orthogonal approach and try to increase the accuracy of NS through a post-processing step that reduces the statistical error in each of the compression factors. To do this, we use IFT to perform Bayesian field inference to reconstruct a continuous and smooth likelihood–prior–volume function given the likelihood contour information from NS. The presented approach has been addressed in Westerkamp et al. (2023). In this paper, we aim to give a deeper introduction into the post-processing and perform further validation.

In the following, we first give a general overview of the methods used in section 2.2. This includes an introduction to NS and its notation, an introduction to IFT and the general explanation of one-dimensional correlated field inference in section 2.2.2, and finally the method for inferring prior volume estimates and a possible implementation of it in section 2.2.3. In section 2.3, we show the according inference results for two validation examples. In particular, we choose a Gaussian likelihood and a spike-and-slab likelihood (Mitchell and Beauchamp, 1988). Finally, we discuss the results, including an analysis of the computational cost, and conclude in section 2.4.

2.2 Methods

2.2.1 Nested sampling algorithm

Using the notation introduced by Skilling (2004) the NS likelihood in Bayesian parameter estimation is denoted via $\mathcal{L}(\theta_M) := \mathcal{P}(d|\theta_M)$ and the prior is denoted by $\pi(\theta_M) := \mathcal{P}(\theta_M)$. The evidence, $\mathcal{Z} := \mathcal{P}(d)$ is calculated accordingly,

$$\mathcal{Z} = \int d\theta_M \mathcal{L}(\theta_M) \pi(\theta_M). \quad (2.3)$$

The idea of NS is to transform this possibly high-dimensional integral in parameter space into a one-dimensional one. Given the prior mass, X , enclosed by some likelihood contour $\mathcal{L}(\theta_M) = L$,

$$X(L) = \int_{\mathcal{L}(\theta_M) > L} d\theta_M \pi(\theta_M), \quad (2.4)$$

we can rewrite eq. (2.3) to a one-dimensional integral,

$$\mathcal{Z} = \int_0^1 dX L(X), \quad (2.5)$$

where $L(X)$ is the likelihood value on the θ -contour that encloses the prior mass X (eq. (2.4)). The underlying algorithm for the calculation of the integral in eq. (2.5) can be summarised as follows: First, n_{live} samples are drawn from the prior $\pi(\theta_M)$, which we call the live points, $\theta_1, \dots, \theta_{n_{\text{live}}}$. For each of these samples the likelihood can be calculated, $d_{L,i} = \mathcal{L}(\theta_i)$. Then, the sample j out of these with the lowest likelihood is added to a new set, called the dead points, $d_L := \{d_{L,j}\}$. A new sample is drawn, restricted to the space of higher likelihood values. This is called likelihood restricted prior sampling (LRPS). Accordingly, we transfer samples from a set of live points to dead points with increasing likelihood while adding new samples to the set of live points for which the likelihood values exceed the highest dead contour. This leads to the condition $d_{L,i} > d_{L,i-1}$. The prior volume under consideration shrinks at each iteration, $X_i < X_{i-1}$, by a compression factor t_i ,

$$X_{i+1} = t_i X_i. \quad (2.6)$$

Under the assumption that the samples are drawn from the prior independently within the highest dead contour, all compression factors, $t = \{t_i\}$, are independent of each other and Beta distributed, $\mathcal{P}(t_i) = \text{Beta}(t_i|1, n_{\text{live}})$. The algorithm stops after n_{iter} iterations. Finally, the set of dead points and estimated prior volumes, defined by eq. (2.6), are used to approximate the evidence using the quadrature rule with the according weights ω_i ,

$$\mathcal{Z} \approx Z = \sum_{i=1}^{n_{\text{iter}}} \omega_i d_{L,i}. \quad (2.7)$$

In this study, we use weights defined by the trapezoidal rule $\omega_i = \frac{1}{2}(X_{i-1} - X_{i+1})$ with $X_0 = 1$ and $X_{n_{\text{iter}}+1} = 0$.

In view of this procedure, NS introduces a statistical uncertainty, since the prior volumes at each iteration, X_i , are not known, but only the distribution of the contraction factors defining the prior volumes is known. In the literature, there are two different approaches for the estimation of the prior volume mentioned. The first one, which we call the statistical approach, samples K chains of compression factors $\{t_i\}_k$ independently. Correspondingly, we can define several sets of prior volumes $\{X_i\}_k$, where for each chain $k = 1, \dots, K$ the prior volume at iteration i is defined by the corresponding sets of contraction factors, $X_{i,k} = \prod_{j=1}^i t_{j,k}$. The result is K samples for the log-evidence using eq. (2.7), which allows us to obtain the mean estimate of the log-evidence and its uncertainty. The second approach, which we call the deterministic approach, instead gives no uncertainty estimate. Here, the mean of the log-compression factors $\langle \ln t_i \rangle_{(t_i)} = -1/n_{\text{live},i}$ is taken as an estimate, as discussed in Feroz et al. (2019). This yields the deterministic prior volume estimation,

$$\bar{X}_i = \prod_{j=1}^i \langle t_j \rangle_{(t_j)} = e^{\ln \left(\prod_{j=1}^i \langle t_j \rangle_{(t_j)} \right)} \approx e^{\sum_{j=1}^i -1/n_{\text{live},j}}, \quad (2.8)$$

and hence $\bar{X}_i \approx e^{-i/n_{\text{live}}}$ if the number of live points remains constant at each iteration (Handley et al., 2015). In other words, the prior volume gets compressed exponentially. In fig. 2.1, we show the likelihood–prior–volume curves generated by NS for a simple Gaussian example, which was introduced by Skilling (2006). In section 2.3.1, the details of this simple Gaussian case are discussed further. The according NS likelihood contours were generated using the software package **anesthetic** (Handley, 2019). Figure 2.1a shows the entire likelihood–prior–volume function generated together with the analytical ground truth. Figure fig. 2.1b shows an enlarged section that is marked in the left panel. In NS each likelihood dead contour, $d_{L,i}$, is accompanied by the estimated prior volume, described by all contraction factors up to the considered iteration. This leads to the corresponding NS likelihood–prior–volume function, defined through either the statistical prior volume estimation, $d_L(X_k)$ for $k = 1, \dots, N$, or the deterministic prior volume estimation, $d_L(\bar{X})$. Both panels of fig. 2.1 show both the statistical and the deterministic likelihood–prior–volume function. The zoomed-in panel also shows the information on the likelihood contours given by NS, which we will use as the only data, d_L for the inference of the likelihood–prior–volume function.

Despite NS being mainly designed to estimate the evidence, it generates posterior samples and their individual probabilities, p_i , at each iteration $i = 1, \dots, n_{\text{iter}}$ of the model parameters θ_M as a by-product as described in Handley et al. (2015),

$$p_i = \frac{\omega_i d_{L,i}}{Z} = \frac{\omega_i d_{L,i}}{\sum_i \omega_i d_{L,i}}. \quad (2.9)$$

These posterior samples can be used to compute the posterior expectation values of any

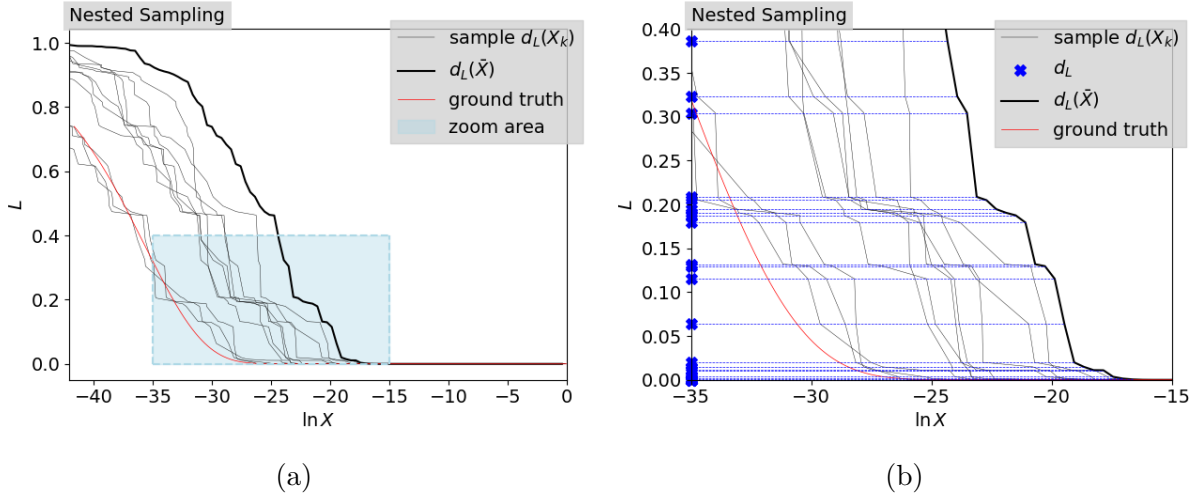


Figure 2.1: Illustration of the nested sampling (NS) output for simple Gaussian example introduced in Skilling (2006) and further elaborated in section 2.3.1 with two live points. Figure 2.1a shows the full NS data generated with `anesthetic` and fig. 2.1b shows a zoomed-in section, which is indicated in fig. 2.1a. The zoomed-in image additionally shows the information of the likelihood dead contours d_L , which we use as data for the Bayesian inference of the prior volumes. In both figures, the samples of likelihood–prior–volume functions defined by prior volume samples, X_k , $d_L(X_k)$, are shown as well as a likelihood–prior–volume function defined by the deterministic NS approach in eq. (2.8), $d_L(\bar{X})$.

function of the model parameters $g(\theta_M)$,

$$\langle g(\theta_M) \rangle_{(\theta_M|d)} = \int d\theta_M g(\theta_M) \mathcal{P}(\theta_M|d) \approx \frac{1}{n_{\text{iter}}} \sum_{i=1}^{n_{\text{iter}}} p_i g(\theta_i). \quad (2.10)$$

One such function and specific quantity of interest is the information gain from the prior to the posterior given by the KL,

$$H := D_{\text{KL}}(\mathcal{P}(\theta_M|d) || \pi(\theta_M)) = \int d\theta_M \mathcal{P}(\theta_M|d) \ln \left(\frac{\mathcal{P}(\theta_M|d)}{\pi(\theta_M)} \right) = \int_0^1 dX \frac{L(X)}{Z} \ln \left(\frac{L(X)}{Z} \right). \quad (2.11)$$

The information gain is an important quantity for estimating the number of steps n_{iter} needed to reach the posterior mass. It is therefore a good measure for determining a termination criterion. In particular, as noted in Skilling (2004, 2012), the posterior set is reached after about $n_{\text{live}} H$ steps and the posterior is passed in $n_{\text{live}} \sqrt{C}$ steps, where C is the number of dimensions. That is, more live points n_{live} provide better sampling of the posterior, but also increase the time to reach the posterior, increasing the overall computation. In addition to the number of live points, n_{live} , the information gain H and the average computational cost for LRPS at each iteration and the average computational cost for evaluating the likelihood have an impact on the total computational cost \mathcal{T} . In particular, \mathcal{T} scales as $\mathcal{O}(n_{\text{live}})$ and $\mathcal{O}(H^2)$ as pointed out in Ashton et al. (2022). Looking at the error ϵ_{NS} for the evidence calculation using NS, we find that, according to Chopin and Robert (2010), it is composed of three components under the assumption that we integrate up to the n_{iter} th iteration using eq. (2.7),

$$\begin{aligned} \epsilon_{\text{NS}} &= \sum_{i=1}^{n_{\text{iter}}} \omega_i d_{L,i} - \int_0^1 L(X) dX \\ &= \underbrace{- \int_0^{X_{n_{\text{iter}}}} L(X) dX}_{\text{truncation error}} + \underbrace{\left(\sum_{i=1}^{n_{\text{iter}}} \omega_i L(X_i) - \int_{X_{n_{\text{iter}}}}^1 L(X) dX \right)}_{\text{numerical integration error}} + \underbrace{\sum_{i=1}^{n_{\text{iter}}} \omega_i (d_{L,i} - L(X_i))}_{\text{stochastic error}}. \end{aligned} \quad (2.12)$$

According to Speagle (2020), the numerical integration error introduced by replacing the integral by the trapezoidal rule is of the order of $\mathcal{O}(1/n_{\text{live}}^2)$ and therefore negligible as the number of live points goes to infinity. The truncation error occurs when we stop at a given maximum iteration n_{iter} . It can be kept small by choosing the stopping criterion wisely. Several approaches have been described in the literature to determine the final iteration n_{iter} , ranging from simultaneously computing the information H to determine the location of the posterior set (Skilling, 2006), to stopping as soon as the LRPS becomes inefficient (Schöniger et al., 2014) or as soon as the expected evidence from the remaining live points compared to the current evidence estimate is less than a user-defined tolerance (Handley et al., 2015; Speagle, 2020; Feroz et al., 2019).

The error we are most interested in and aim to minimise is the stochastic error introduced by the unknown prior volumes. This error is of the order of $\mathcal{O}(n_{\text{live}}^{-\frac{1}{2}})$ (Chopin and Robert, 2010), or as derived in Skilling (2004), proportional to $\sqrt{H/n_{\text{live}}}$, and dominates the evidence approximation error. As a result, the stochastic error trends to zero as the number of live points goes to infinity, but at the same moment the computational cost trends to infinity. Thus, there is a trade-off between the accuracy of the evidence calculation and the computation time. In the following, we will present a post-processing step for NS that aims to reduce either the error in the evidence calculation or the time complexity, depending on the measure of interest.

2.2.2 Gaussian processes and information field theory

We use information field theory (IFT) (Enßlin, 2019) for the joint reconstruction of the continuous likelihood–prior–volume function and the discrete set of prior volumes. IFT focuses on Bayesian field inference, or in other words, on the reconstruction of a continuous field from a discrete dataset. Here, we consider the likelihood–prior–volume function to be a one-dimensional field with an infinite number of degrees of freedom, which is to be reconstructed from a finite set of likelihood dead contours, d_L . Thus, the inference problem is underconstrained, and we need prior knowledge of the likelihood–prior–volume function and the prior volumes to obtain the posterior. We call this prior probability distribution the joint reconstruction prior $\mathcal{P}(L(X), t)$ to avoid confusion with the prior volumes of NS. The reconstruction likelihood $\mathcal{P}(d_L|L(X), t)$ is then the probability of the measured likelihood dead contours given the likelihood–prior–volume function. We combine the information on the reconstruction prior and the reconstruction likelihood in Bayes’ theorem (eq. (2.1)) to reconstruct the posterior probability of the field $L(X)$, which we call the reconstruction posterior $\mathcal{P}(L(X), t|d_L)$. This allows us to obtain any a posteriori measure of interest, like for example the mean and variance of the likelihood–prior–volume function. In section 2.2.3, we will introduce the explicit reconstruction likelihood and prior models for the here introduced inference. Here, we focus on using IFT and its software package NIFTy (Arras et al., 2019) to implement a generative prior for the likelihood–prior–volume function, considering its correlation structure.

Since the prior model is based on Gaussian processes, this section describes Gaussian processes from the IFT perspective. Specifically, we introduce a generative model for Gaussian processes with variable correlation structure. In other words, the aim is to generate a field, τ , as a Gaussian process $\mathcal{N}(\tau, T)$ with an unknown covariance T . This implementation, which is desirable in many cases, ensures the smoothness of the likelihood–prior–volume function. We use this information in the presented algorithm to improve the accuracy of the evidence calculation. In the following sections, we will discuss the smoothness assumption and its positive and negative consequences in more detail. One approach for the implementation of a non-parametric model for fields with unknown correlation structure, was introduced in Arras et al. (2022). Analogously, we call this model the correlated field model, which suggests that not only the realization of the field itself, τ , is

learned, but also the underlying correlation structure T . Here, we consider the simplest case of a one-dimensional correlated field. The correlated field is implemented as a generative process using the reparametrization trick introduced in Kingma et al. (2015), $\tau = A\xi_\tau$ with $T = AA^\dagger$ and $\mathcal{P}(\xi_\tau) = \mathcal{N}(\xi_\tau, \mathbb{I})$. We can separate the field realization from the field correlation structure using this basis transformation. This means that the new coordinates, ξ_τ , have the same dimension as τ , but are a priori uncorrelated. Assuming statistical homogeneity and isotropy, the covariance T is fully defined by its power spectrum $p_T(|q|)$ via the Wiener–Khinchin theorem in Fourier space (Khinchine, 1934),

$$A_{qq'} = (FAF^\dagger)_{qq'} = 2\pi\delta(q - q')\sqrt{p_T(|q|)}, \quad (2.13)$$

where F is the Fourier transform and $\sqrt{p_T(|q|)}$ is the amplitude spectrum. The aim is to infer the power spectrum non-parametrically. This is achieved by building a model of the power spectrum where each hyper-parameter is described by a Gaussian or log-normal prior with a given mean and standard deviation. Each hyper-parameter, and thus the entire power spectrum, is learned during inference. More specifically, the amplitude spectrum is implemented as an integrated Wiener process, a general continuous process, on the logarithmic scale $l = \log(|q|)$ for $q \neq 0$,

$$\sqrt{p_T(l)} \propto e^{\gamma(l)}, \quad \frac{d^2\gamma}{dl^2} = \eta\xi_W(l), \quad \mathcal{P}(\xi_W) = \mathcal{N}(\xi_W, \mathbb{I}). \quad (2.14)$$

The integration gives

$$\gamma(l) = ml + \eta \int_{l_0}^l \int_{l_0}^{l'} \xi_W(l'') dl' dl''. \quad (2.15)$$

Here, l_0 is the first mode greater than zero and m defines the slope of the integrated Wiener process, i.e. it is the slope of the amplitude spectrum on a double logarithmic scale. The parameter η is called flexibility because it controls the total variance of the integrated Wiener process. In addition to these parameters, m and η , which essentially determine the shape of the power spectrum, the total offset of the correlated field defined by the zero mode and another hyper-parameter, called the fluctuations, a , which specifies the total fluctuations of the non-zero modes,

$$\sqrt{p_T(l)} = a \frac{e^{\gamma(l)}}{\left(\int_{l_0}^l e^{2\gamma(l')} dl' \right)^{\frac{1}{2}}}, \quad (2.16)$$

are introduced. The expression in the denominator normalizes, so that the meaning of a as the fluctuation amplitude is invariant under the change of γ , which determines on which Fourier scales the fluctuations appear. Overall, this gives a generative model for a Gaussian random field with unknown covariance.

The effect of changing a , m , and η is shown in fig. 2.2. It shows a reference power spectrum and corresponding sample field realizations, together with the power spectrum

and field realizations for a changed mean of one of the hyper-parameters. The variances of the hyper-parameters are the same for all cases and are kept small in order to illustrate the effect of the hyper-parameters on the correlated field more effectively. The specific means and variances of the hyper-parameters are listed in table 2.1. For the reconstruction itself, we keep the prior wide, which means that we take higher values for the standard deviations of each parameter to allow for flexibility of the model.

IFT performs Bayesian field inference to infer the posterior for a continuous field given some data d_τ . The exact relationship between the correlated field, τ , and the likelihood–prior–volume function is discussed in section 2.2.3. Thereby, the posterior probability $\mathcal{P}(\tau|d_\tau)$ is approximated by a simpler posterior distribution $\mathcal{Q}(\tau|d_\tau)$ using variational inference (VI). The approximation is done by minimizing the cross-entropy term of the KL between the actual posterior and its approximation $D_{\text{KL}}(\mathcal{Q}(\tau|d_\tau)||\mathcal{P}(\tau|d_\tau))$. In particular, we use the geoVI introduced by Frank et al. (2021). The geoVI algorithm optimizes the cross-entropy of the KL with respect to a non-linear normalizing coordinate transformation that maps the posterior onto a standard Gaussian. This allows it to approximate non-Gaussian posteriors. All numerics related to IFT are implemented in the corresponding software package NIFTy (Arras et al., 2019).

Table 2.1: Hyper-parameters for correlated field samples shown in fig. 2.2. The reference parameters are denoted by an index **r**. The other indices correspond to the labels of the sub-figures (**a**, **b**, **c**). Modified hyper-parameter means with respect to the reference field are marked in blue.

	$\eta_{\mathbf{r}}$	$m_{\mathbf{r}}$	$a_{\mathbf{r}}$	$\eta_{\mathbf{a}}$	$m_{\mathbf{a}}$	$a_{\mathbf{a}}$	$\eta_{\mathbf{b}}$	$m_{\mathbf{b}}$	$a_{\mathbf{b}}$	$\eta_{\mathbf{c}}$	$m_{\mathbf{c}}$	$a_{\mathbf{c}}$
Mean	0.5	−6	1.0	0.5	−6	3.0	0.5	−2	1.0	10.0	−6	1.0
Std	0.5	10^{-16}	0.5	0.5	10^{-16}	0.5	0.5	10^{-16}	0.5	0.5	10^{-16}	0.5

2.2.3 Bayesian inference of the likelihood–prior–volume function

As noted in Westerkamp et al. (2023), we use the smoothness assumption for the likelihood–prior–volume curve to improve the evidence calculation in NS. The description of the algorithm is given below using the simple Gaussian example introduced by Skilling (2006) for illustration. The full information from NS for this case is presented in fig. 2.1a. As mentioned in section 2.2.1, NS generates data on the likelihood dead contours d_L . These data points are marked in fig. 2.1 on the likelihood axis. However, to compute the evidence we would need the likelihood–prior–volume function, including information on the prior volumes. In the following, we aim to give an algorithm that aims to improve the overall estimate on the prior volume and thereby reduces the uncertainty in the evidence.

We propose to jointly infer the likelihood–prior–volume function $L(X)$ and the prior volumes $\{X_i\}$ at each iteration $i = 1, \dots, n_{\text{iter}}$ using Bayesian field inference as described in section 2.2.2. For the data, we only take into account the information we obtain from NS about the probability of dead contours, d_L . The composition of the reconstruction is

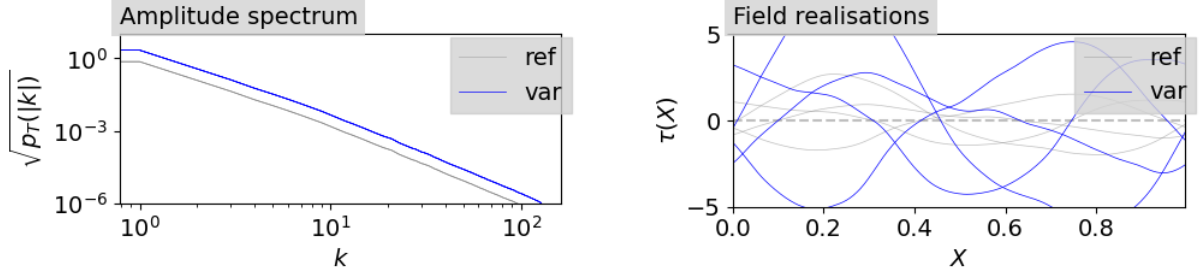
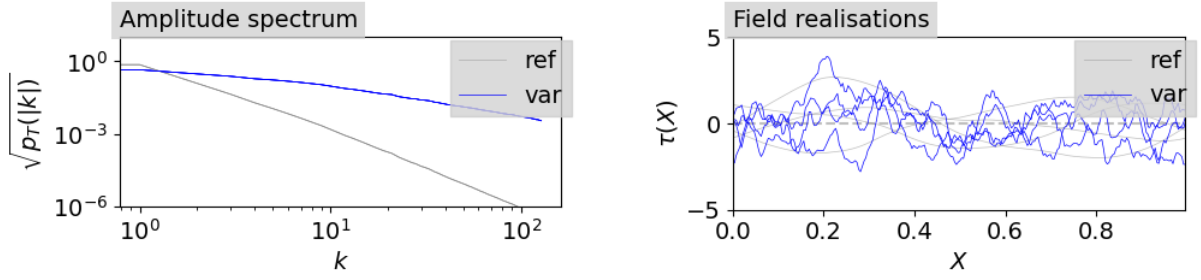
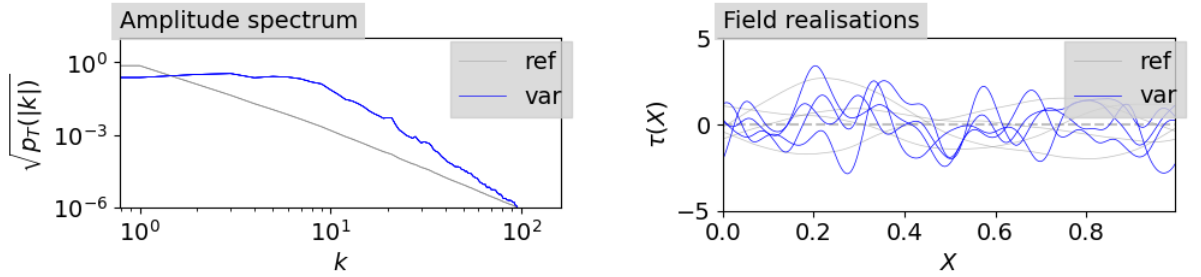
(a) Change in a :(b) Change in m :(c) Change in η :

Figure 2.2: Visualization of the effect of changes in the mean value for one of the hyper-parameters of the power spectrum model. The changed amplitude spectrum is shown on the left side and the according influence on the field realization is shown on the right side by comparison of a reference correlated field (ref) and a variation of one of its hyper-parameters (var). Which hyper-parameter is changed and how it is changed in comparison to the reference is denoted in table 2.1.

shown in fig. 2.3a. As ingredients for Bayes theorem, we use the joint reconstruction prior model as well as the reconstruction likelihood model, which in our case is defined fully by the data. Figure 2.3a shows prior samples for the likelihood–prior–volume function described by the correlated field to be learned. The data obtained by NS for the likelihood dead contours, which then describes the reconstruction likelihood, are shown in fig. 2.3b. Merging these two models and approximating the reconstruction posterior with VI yields posterior samples of the likelihood–prior–volume function and the set of prior volumes. Figure 2.3c shows the computed mean and uncertainty for the reconstructed posterior likelihood–prior–volume function, as well as the function obtained by pure NS and the analytic ground truth. In fig. 2.3d, a zoomed area on the reconstructed likelihood–prior–volume function is shown, in order to facilitate comparison with fig. 2.1b.

Below, we describe a method to enforce the smoothness assumption on the likelihood–prior–volume curve using the correlated field model introduced in section 2.2.2. In appendix A.1, we present an alternative approach that requires no δ -function approximation; however, this method is currently only applicable for a MAP estimate and not for VI. The here presented approach implements the smoothness by describing the derivative of the likelihood by the prior volume as a log-normal process, which can be achieved by using a correlated field as described in section 2.2.2. By using a log-normal rather than a Gaussian process for the derivative of the logarithmic likelihood–prior–volume relation, we ensure that the likelihood–prior–volume function is monotonic. As a consequence of the correlated field model described above, it would be desirable that

$$-e^{-\tau(\ln X)} = \frac{d \ln L}{d \ln X} \approx \text{const}, \quad (2.17)$$

with τ drawn from a Gaussian process. However, it can be seen that for the simplest case, a Gaussian likelihood model, this assumption is not fulfilled. Accordingly, we introduce a reparametrization f that maps $\ln L$ such that we find a damped log-normal process for $\frac{d \ln L}{d \ln X}$,

$$\frac{df_{\ln L}}{d \ln X} := \frac{df(\ln L)}{d \ln X} = \frac{1}{\ln L_{\max} - \ln L} \frac{d \ln L}{d \ln X} = -e^{-\tau(\ln X)}. \quad (2.18)$$

This way, a constant τ perfectly captures the Gaussian case, and non-Gaussianity is absorbed in excitations of τ around this constant, as described in section 2.2.2. The derivation of the reparametrization for the Gaussian case is given in appendix A.2. Appendix A.3 shows how to calculate the corresponding L_{\max} according to Handley and Lemos (2019) if it is not known analytically. The joint reconstruction prior for the reparametrized likelihood–prior–volume function $f_{\ln L}$ and the contraction factors t is fully defined via the joint prior $\mathcal{P}(\tau, t)$,

$$\mathcal{P}(f_{\ln L}, t) = \mathcal{P}(f_{\ln L}|t)\mathcal{P}(t) = \mathcal{N}(\tau, T) \prod_{i=1}^{n_{\text{iter}}} \text{Beta}(t_i|1, n_{\text{live},i}) = \mathcal{P}(\tau, t). \quad (2.19)$$

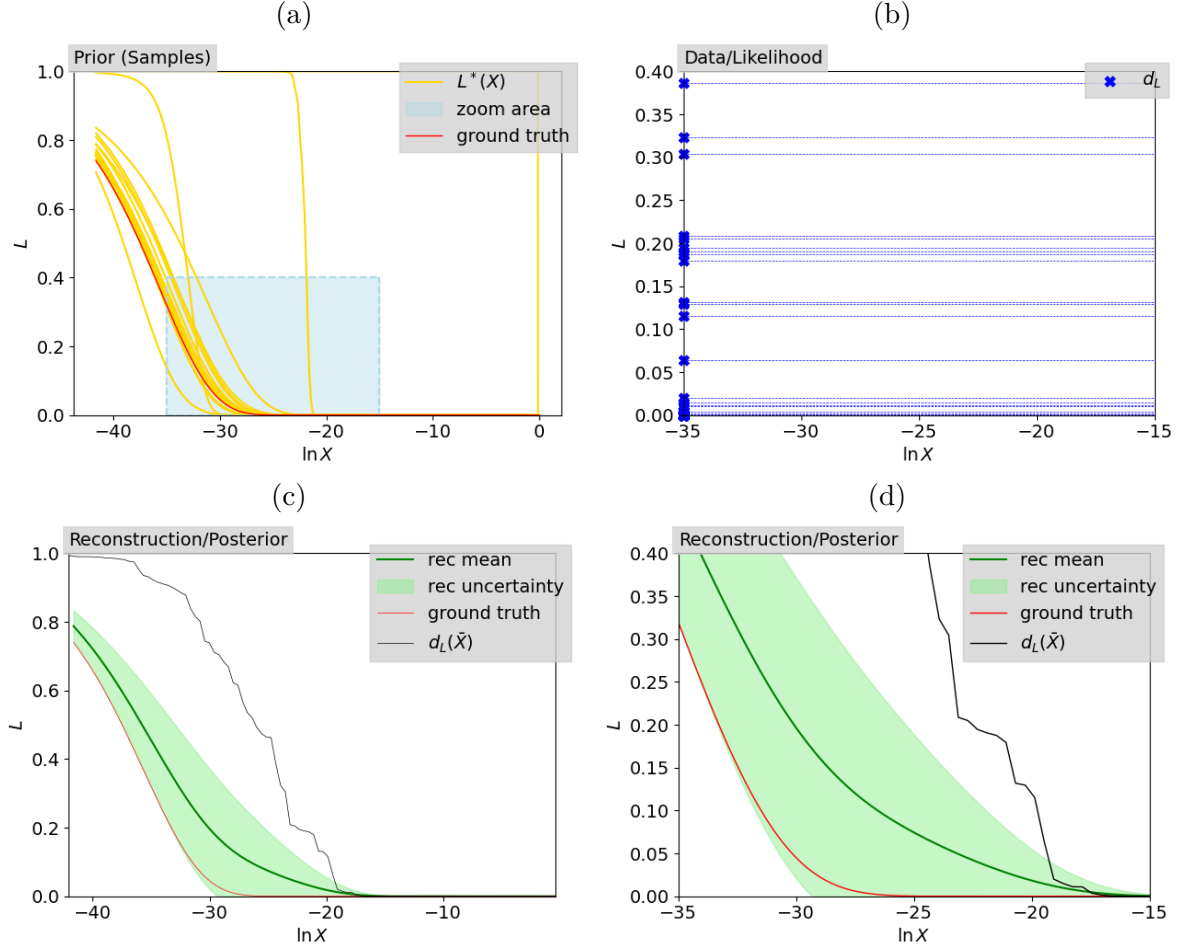


Figure 2.3: Illustration of the Bayesian field inference process for the simple Gaussian, which is further elaborated in section 2.3.1 for two live points. The prior samples, the data used, and the final reconstruction compared to the NS approach and the ground truth are shown. Figure 2.3a: Prior samples for the likelihood-prior-volume function ($L^*(X)$) in yellow together with the ground truth. A zoomed-in area is marked beside it (the same area as in fig. 2.1a) which is taken to zoom into the data in fig. 2.3b and the reconstruction in fig. 2.3d. Figure 2.3b: Data on likelihood dead contours for the given zoom area. Figures 2.3c, 2.3d: Reconstruction mean (rec mean) of the likelihood-prior-volume function and the associated uncertainty, defined via the one sigma contours (rec uncertainty), which is zoomed in on in fig. 2.3d, and the full image is presented in fig. 2.3c. The result for the likelihood-prior-volume function for the deterministic NS approach is shown ($d_L(\bar{X})$), which is the same as in fig. 2.1b, and the analytic likelihood-prior-volume-function (ground truth).

The reconstruction likelihood is described by the solution of the differential equation in eq. (2.18), which can be written as a function of the set of contraction factors,

$$f_{\ln L}(t) = f_{\ln L}(\ln X_j = \sum_{i=1}^j t_i) = f_{\ln L}(0) - \int_0^{\sum_{i=1}^j t_i} e^{-\tau(z)} dz. \quad (2.20)$$

This leads to a likelihood model, which is a δ -function, which we approximate by a Gaussian with a small chosen variance σ_δ ,

$$\mathcal{P}(f(\ln d_L)|\tau, t) = \delta(f(\ln d_L) - f_{\ln L}(t)) \approx \mathcal{N}(f(\ln d_L) - f_{\ln L}(t), \sigma_\delta). \quad (2.21)$$

The result is the joint reconstruction posterior,

$$\mathcal{P}(\tau, t|f(\ln d_L)) \propto \mathcal{P}(f(\ln d_L)|\tau, t)\mathcal{P}(\tau, t). \quad (2.22)$$

The posterior is approximated using geoVI as described in section 2.2.2, slowly increasing the number of samples from iteration to iteration. We choose $\sigma_\delta = 0.1 \min(\text{dist}(f(\ln d_L)))$ to ensure that the likelihood does not allow for the exchange of two data points or even non-monotonicity. Finally, the prior of τ prefers a constant flat τ without excitation, corresponding to a linear relation between $f_{\ln L}$ and $\ln X$, corresponding to a Gaussian likelihood in NS. However, the prior of t prefers certain distances between the prior volumes given by the Beta distribution. The likelihood ensures that the reconstructed function $f_{\ln L}$ evaluated at the reconstructed prior volumes X matches the values of $f(\ln d_L)$. This means that, for example, if there is a jump in $f(\ln d_L)$, $\mathcal{P}(\tau)$ prefers a corresponding jump in $\ln X$ to ensure smoothness and to avoid deviations from the linear reparametrized likelihood–prior–volume relation, while $\mathcal{P}(t)$ on average prefers an increase in $\ln X$ defined by the Beta distribution.

2.3 Results and analysis

For validation, we consider two cases: a simple Gaussian case as described in Skilling (2006) and a spike-and-slab likelihood as introduced in Mitchell and Beauchamp (1988). The according data on the likelihood live and dead contours are generated using the **anesthetic** package by Handley (2019). These test cases are valuable for checking the consistency of the presented method, as they allow the analytical calculation of evidence.

2.3.1 Gaussian case

As a first validation test case, we use a zero-centered Gaussian likelihood,

$$L(\theta) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{with} \quad r^2 = \sum_{i=1}^C \theta_i^2, \quad \theta = \{\theta_i\}_{i=1,\dots,C}, \quad (2.23)$$

Table 2.2: Inferred (IFT) and NS (NS stat: statistical, NS det: deterministic) results for the computed Gaussian log-evidence, represented by the mean and the according standard deviation for $n_{\text{live}} \in \{2, 10, 1000\}$ given K sample chains. The ground truth is $\ln Z = -30.87$. The histograms of the distributions of the evidences are shown in fig. 2.4.

n_{live}	K	Mean			Standard Deviation	
		IFT	NS Stat	NS Det	IFT	NS Stat
2	120	-28.51	-25.11	-20.90	2.98	3.23
10	120	-30.56	-30.44	-29.37	0.24	1.6
1000	120	-31.01	-30.52	-30.51	0.61	0.17

in $C = 10$ dimensions with as variance $\sigma = 0.02$. As a prior, we use, in analogy to Skilling (2006), a flat prior on the unit sphere

$$\pi(\theta) = \frac{C/2!}{\pi^{C/2}} \quad \text{with } r < 1. \quad (2.24)$$

The evidence given the probability and prior above can be calculated analytically to be

$$\mathcal{Z} = \int_{-\infty}^{\infty} d\theta^C L(\theta)\pi(\theta) = \frac{(C/2)!}{(2\sigma^2)^{C/2}}. \quad (2.25)$$

The definition of the prior mass in C dimensions is given by $X = r^C$, which allows us to compute the ground truth of the likelihood–prior–volume function,

$$L(X) = \exp\left(\frac{-X^{2/C}}{2\sigma^2}\right). \quad (2.26)$$

Figure 2.4 shows the ground-truth likelihood–prior–volume function together with samples of the likelihood–prior–volume function defined by the reconstructed prior volumes or the statistical approximated prior volumes from NS for a constant number of live points, $n_{\text{live}} \in \{2, 10, 1000\}$. Also shown is the likelihood–prior–volume function for the deterministic NS approach using eq. (2.4). As described in section 2.2.1, it can be seen that the standard deviation for the NS approach decreases as the number of live points increases. For each of these prior volume estimation approaches, the statistical NS, the deterministic NS or the IFT based, we calculate the log-evidence using the weighted sum in eq. (2.7). This gives us sample sets of evidence for the statistical NS and IFT approaches, which are plotted as a histogram in fig. 2.4 together with the analytical ground truth and the deterministic NS approach. The corresponding results for each of the approaches for the mean and, where applicable, the standard deviation are given in table 2.2. Further discussion of the results can be found in section 2.4.

2.3.2 Spike-and-slab case

As a next step we consider a non-Gaussian test case for validation. In particular, we will look at a spike-and-slab likelihood known from Bayesian variable selection (Mitchell and

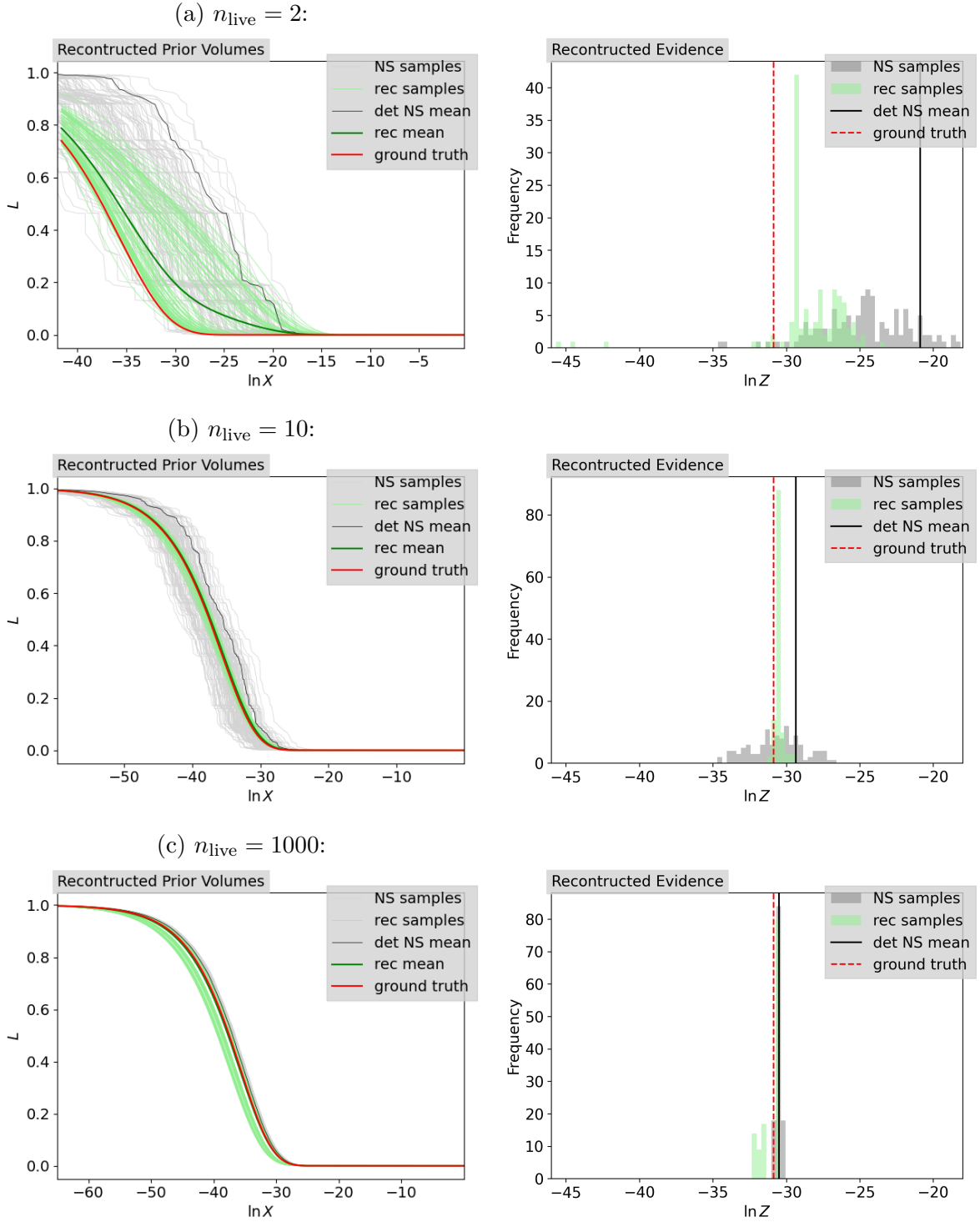


Figure 2.4: Reconstruction results for the Gaussian prior volumes accompanied by the likelihood contours given by NS on the left and the computed log-evidence on the right for $n_{\text{live}} \in \{2, 10, 1000\}$ from top to bottom. The inferred posterior samples (rec samples) are shown together with their mean (rec mean) and compared with the corresponding statistical (NS samples) and deterministic (det NS mean) NS results and the ground truth.

Beauchamp, 1988), which is the sum of a zero-centered spike and a broad Gaussian background. This leads to an abrupt change in the prior volume X with increasing likelihood. The corresponding likelihood is given by,

$$L(\theta) = A \exp\left(-\frac{r^2}{2\sigma_1^2}\right) + (1 - A) \exp\left(-\frac{r^2}{2\sigma_2^2}\right) \quad \text{with} \quad r^2 = \sum_{i=1}^C \theta_i^2. \quad (2.27)$$

Again, we choose a flat prior as described in eq. (2.24). Accordingly, we are able to calculate the evidence analytically, which gives us a good point for comparison,

$$\mathcal{Z} = \int_{-\infty}^{\infty} d\theta^C L(\theta) \pi(\theta) = C/2! \left(a(2\sigma_1^2)^{C/2} + (1 - a)(2\sigma_2^2)^{C/2} \right). \quad (2.28)$$

Just as in section 2.3.1, we can obtain the analytic likelihood–prior–volume function given the prior volume $X = r^C$,

$$L(X) = A \exp\left(\frac{-X^{2/C}}{2\sigma_1^2}\right) + (1 - A) \exp\left(\frac{-X^{2/C}}{2\sigma_2^2}\right). \quad (2.29)$$

The parameters of the spike-and-slab likelihood under consideration are denoted in table 2.3.

Table 2.3: Parameters for the spike-and-slab likelihood described in eq. (2.29).

Parameter	Meaning	Value
C	number of dimensions	10
A	relative weight of Gaussians	0.5
σ_1	std of Gaussian weighted by a	0.1
σ_2	std of Gaussian weighted by $(1 - a)$	0.02

Figure 2.5 shows the ground truth of the likelihood–prior–volume function (eq. (2.29)) and evidence (eq. (2.28)) together with the corresponding samples and mean (eq. (2.4)) given by the corresponding NS runs for $n_{\text{live}} \in \{2, 10, 1000\}$. We infer the likelihood–prior–volume function and the set of prior volumes jointly using the inference algorithm described in section 2.2.3. As a result, we obtain a set of posterior samples on the likelihood–prior–volume functions and the prior volumes, which leads to a set of posterior samples of the evidence using eq. (2.7). The posterior samples for the likelihood contours as a function of the reconstructed prior volumes (rec samples) as well as their mean (rec mean) are shown on the left side of fig. 2.5. The computed corresponding evidence is shown on the right side of fig. 2.5. For comparison, additionally the results for the statistical NS approach

and the deterministic NS approach as well as the ground truth are shown. The computed mean evidences for the sample sets for the IFT and the NS approach are listed in table 2.4 together with the corresponding standard deviation. A further discussion of the results is given in section 2.4.

Table 2.4: Inferred (IFT) and NS (NS stat: statistical, NS det: deterministic) results for the computed spike-and-slab log-evidence, represented by the mean and the according standard deviation for $n_{\text{live}} \in \{2, 10, 1000\}$ given K sample chains. The ground truth is $\ln Z = -15.47$. The histograms of the distributions of the evidences are shown in fig. 2.5.

n_{live}	K	Mean			Standard Deviation	
		IFT	NS Stat	NS Det	IFT	NS Stat
2	40	-17.26	-13.00	-11.21	1.67	1.91
10	40	-16.03	-14.62	-14.01	0.78	0.80
1000	20	-18.35	-15.49	-15.50	0.01	0.10

2.4 Discussion

In NS, the statistical error in one of the contraction factors, t_i , affects each upcoming prior volume according to eq. (2.6). For this reason, it has a major impact on the calculation of the logarithmic evidence. However, the propagation of error does not occur on the likelihood contour information, which is assumed to be accurate, but only on the prior volume estimates. We use additional knowledge to be able to assign an improved estimate of the prior volume to the corresponding likelihood contour. The assumption we take into account is the smoothness of the likelihood–prior–volume function, which is valid for a large set of problems. Of course, if a problem is considered where the likelihood–prior–volume function is not smooth, this post-processing algorithm will not be applicable and the inference of the prior volumes could even make the result worse. One extreme example, which was discussed in Fowlie et al. (2021), is the wedding cake likelihood. Furthermore, the inference algorithm has its limits when dealing with likelihood plateaus. The reason for this limitation is that we model the derivative of the likelihood–prior–volume function as a modified log-normal process. A zero slope would therefore correspond to an infinite excitation of the correlated field τ . A way to deal with this problem was suggested by Schittenhelm and Wacker (2021). Accordingly, one could split the dataset into several parts and perform the inference solely on the strictly negative monotonic regions of the likelihood–prior–volume functions. The same applies to the use of nested sampling to compute the evidence for multimodal likelihoods. In Feroz and Hobson (2008), three different approaches to LRPS for multimodal likelihoods are presented; as long as the final likelihood–prior–volume function would be given by a smooth function that could be partitioned around possible plateaus, the algorithm outlined here is well-suited for the task.

In terms of computational cost, our overall goal is to define a post-processing algorithm

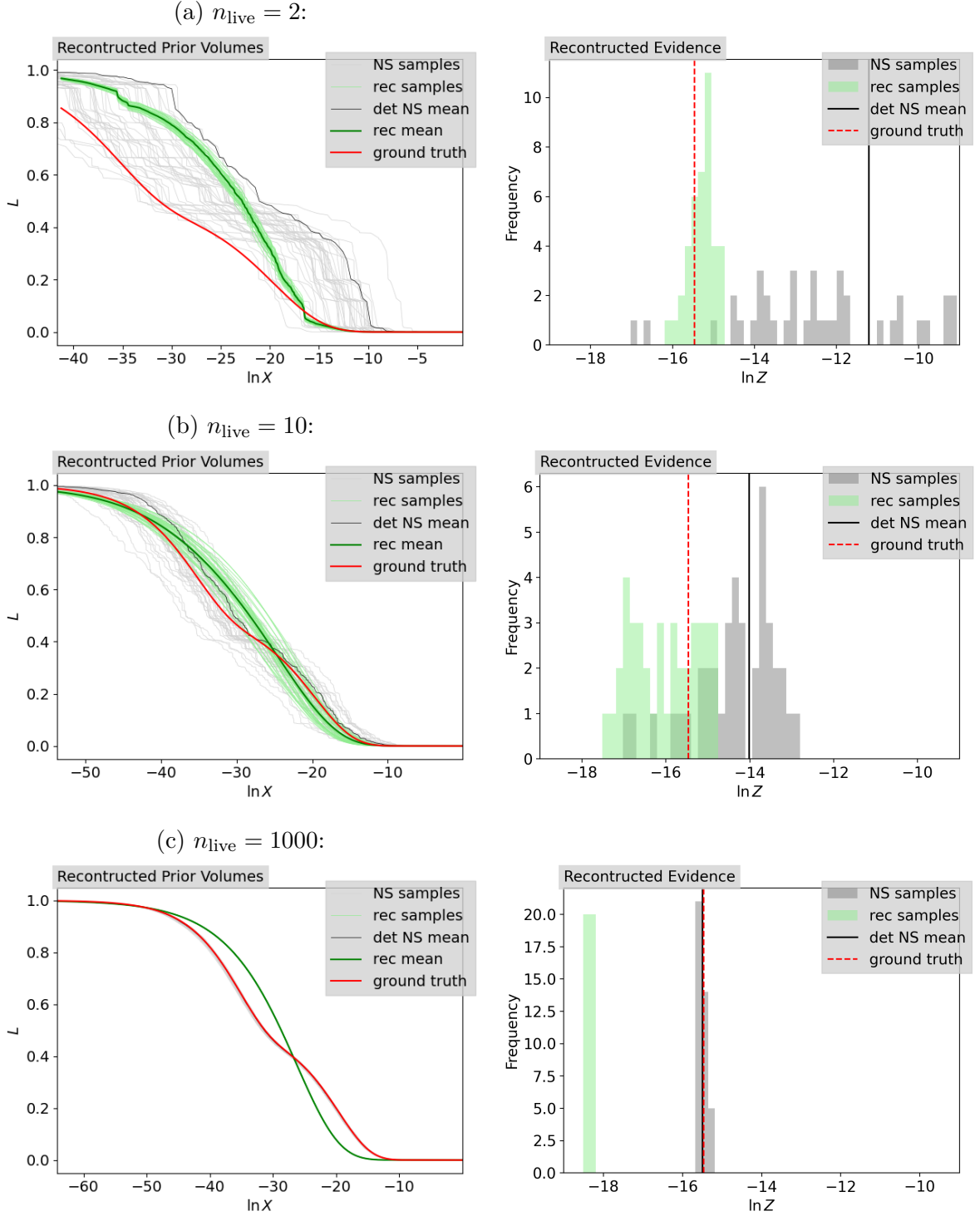


Figure 2.5: Reconstruction results for the spike-and-slab prior volumes accompanied by the likelihood contours given by NS on the left and the computed log-evidence on the right for $n_{\text{live}} \in \{2, 10, 1000\}$ from top to bottom. The inferred posterior samples (rec samples) are shown together with their mean (rec mean) and compared with the corresponding statistical (NS samples) and deterministic (det NS mean) NS results and the ground truth.

whose computational cost is independent of the number of live points, n_{live} . As described in section 2.2.1, the computational cost of a NS run is proportional to the number of live points n_{live} and the KL, H , while the error in $\ln Z$ is proportional to \sqrt{H} and anti-proportional to $\sqrt{n_{\text{live}}}$. Therefore, a large shrinkage from prior to posterior increases both the error and the computation time, while the error can be reduced by using more live points, leading to an increase in computation time. Accordingly, if the inference described here reduces the error by a factor of ϵ_{NS} , then we could assume that the same result can be achieved using just the standard NS run with more live points. In particular, the number of live points would have to be increased by ϵ_{NS}^2 , which would also increase the computational complexity by a factor of ϵ_{NS}^2 , i.e., $\mathcal{T}_\epsilon = \epsilon_{\text{NS}}^2 \mathcal{T}$. In contrast, the post-processing with IFT adds a constant, live point independent computational complexity \mathcal{T}_{IFT} to the original computational effort \mathcal{T} in order to reduce the error by ϵ , i.e., $\mathcal{T}_{\epsilon, \text{IFT}} = \mathcal{T} + \mathcal{T}_{\text{IFT}}$. Therefore, in cases where \mathcal{T} is high, e.g., when the KL is high, or in particular when the evaluation of the likelihood or the sampling from the restricted prior requires a lot of time, it makes sense to consider the inference of the likelihood–prior–volume function instead of the addition of further live points.

If we examine the reconstructed likelihood–prior–volume curves shown in section 2.3, we find that the prior volumes for the spike-and-slab likelihood show an abrupt change in the prior volume in the interval $\ln X \in [-20, -30]$. In fact, this quick change in the prior volume corresponds to small slope of the likelihood–prior–volume curve in this interval, which would need to be modelled by a large value of τ . Looking at the prior generative model shown in section 2.2.2, this particular case is difficult to model, as the correlated field would need to be a straight line with a peak in this interval, which it is hardly forced to by the relatively few data points given for $n_{\text{live}} = 2$ or $n_{\text{live}} = 10$. We expect this to look better for examples that have several changes in the slope of the curve. In any case, looking at the results in table 2.2 and table 2.4, we see that the reconstruction works quite well during validation for the Gaussian and spike-and-slab examples. However, it is noticeable that the benefit of using post-processing decreases as the number of live points increases. This is expected because the likelihood–prior–volume function generated by NS itself becomes smoother as the number of live points increases. Nevertheless, we can see an increase in the accuracy of the result for the log-evidence up to one hundred live points in terms of a decreased standard deviation. When the number of live points increases even further, we find that the final σ_δ , defined by the minimum distance between two adjacent nested likelihood contours, becomes very small. This can especially be seen in the case of the spike-and-slab likelihood for one thousand live points, where $\sigma_\delta = 2.26 \times 10^{-10}$. As a result, the reconstruction using the Gaussian approximation of the δ -function becomes numerically unstable, in the sense of a high reduced χ^2 value between the data and the reconstruction, allowing the reconstruction to be performed with fewer sample chains and obtaining a worse estimate of the log evidence. Therefore, the likelihood–prior–volume function inference proposed here is not applicable to a large number of live points. Instead, it could be used to improve the log-evidence calculation in scenarios where only a small number of live points are feasible, e.g., due to high LRPS costs. The applicability with respect to the reduced χ^2 and the inherent uncertainty given by the algorithm should

therefore be checked by the user.

Thus, this post-processing method leaves some room for future work. First of all, it would be desirable to apply it to a problem with a large number of live points. Two different approaches could be considered. First, one could split the dataset into parts and thus generate NS datasets with a smaller number of live points and thus a larger distance between adjacent nested likelihood contours (Skilling, 2006). Second, one could think of ways to avoid the Gaussian approximation of the δ -function. One alternative approach that does not require the Gaussian approximation of the δ -function is described in the appendix A.1. Using this approach for VI is left for future work. Moreover, one could think of further assumptions besides the smoothness assumption and include them in the inference to allow for a wider range of likelihoods, such as non-smooth ones or a specific set of likelihoods.

In conclusion, we have presented a post-processing step for NS that uses a smoothness assumption to infer the likelihood–prior–volume function, providing an estimate of the inherent uncertainty in the reconstruction due to the stochastic approach. This aims to reduce the statistical error in the evidence computation introduced by the unknown prior volumes at each iteration step. Since this post-processing can deal with a varying number of live points, it is applicable to advanced NS methods such as dynamic NS. Finally, some work needs to be done to apply it to problems with a larger number of live points. However, the method presented here should provide a first approach to improving prior volume estimation, which remains one of the main challenges in NS.

2.5 Acknowledgements

Margret Westerkamp acknowledges support for this research through the project Universal Bayesian Imaging Kit (UBIK, Förderkennzeichen 500O2103) funded by the Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR). Jakob Roth acknowledges financial support by the German Federal Ministry of Education and Research (BMBF) under grant 05A20W01 (Verbundprojekt D-MeerKAT) and grant 05A23W01 (Verbundprojekt D-MeerKAT III). Philipp Frank acknowledges funding through the German Federal Ministry of Education and Research for the project ErUM-IFT: Informationsfeldtheorie für Experimente an Großforschungsanlagen (Förderkennzeichen: 05D23EO1). This research was supported by the Munich Institute for Astro-, Particle and BioPhysics (MIAPbP) which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy- EXC-2094-390783311.

Chapter 3

First spatio-spectral Bayesian imaging of SN1006 in X-rays

*The following chapter has been published in **Astronomy & Astrophysics** with me as the first author (Westerkamp et al., 2024a). My contributions to this project are the implementation of the likelihood models, the prior models, the inference script, and an extensive diagnostic pipeline. I also performed the hyper-parameter search and computational analysis, and wrote the manuscript. The project grew out of a close collaboration between Vincent Eberle and myself. Vincent Eberle and I implemented the initial inference pipeline for SN1006 using multiple Chandra observations and a spatial prior model. This was greatly facilitated by the efforts of Julia Stadler, who implemented a data loading pipeline for Chandra observations. In exchange with Lukas Platz, I implemented the first spatio-spectral prior model for the X-ray sky in NIFTy. To speed up the reconstruction with the spatio-spectral prior, I implemented a so-called transition model based on fruitful discussions with Jakob Knollmüller and Philipp Arras. This work also benefited from discussions with Torsten Enßlin and Philipp Frank on validation, diagnostics, and feedback on the methodological approach, as well as from the scientific exchange with Matteo Guardiani. All authors read, commented and approved the final manuscript.*

For consistency within this thesis, some of the parameters have been renamed and some of the figures have been adapted according to the layout.

Supernovae (SNe) are an important source of energy in the interstellar medium. Young remnants of supernovae exhibit peak emission in the X-ray region, making them interesting objects for X-ray observations. In particular, the supernova remnant (SNR) SN1006 is of great interest due to its historical record, proximity, and brightness. Thus, it has been studied with a number of X-ray telescopes. Improving X-ray imaging of this and other remnants is an important but challenging task, as it often requires multiple observations with different instrument responses to image the entire object. Here, we use Chandra observations to demonstrate the capabilities of Bayesian image reconstruction using information field theory (IFT). Our objective is to reconstruct denoised, deconvolved, and spatio-spectral resolved images from X-ray observations and to decompose the emission into different morphologies, namely, diffuse and point-like. Further, we aim to fuse data from different detectors and pointings into a mosaic and quantify the uncertainty of our result. By utilizing prior knowledge on the spatial and spectral correlation structure of the diffuse emission and point sources, this method allows for the effective decomposition of the signal into these two components. In order to accelerate the imaging process, we introduced a multi-step approach, in which the spatial reconstruction obtained for a single energy range is used to derive an informed starting point for the full spatio-spectral reconstruction.

We applied this method to 11 Chandra observations of SN1006 from 2008 and 2012, providing a detailed, denoised, and decomposed view of the remnant. In particular, the separated view of the diffuse emission ought to provide new insights into the complex, small-scale structures in the center of the remnant and at the shock front profiles. For example, our analysis reveals sharp X-ray flux increases by up to two orders of magnitude at the shock fronts of SN1006.

3.1 Introduction

In the year 1006, observers on Earth were able to see the light of a bright "new star," which eventually faded after a few months. This observation is now attributed to a Type 1a supernova (SN1a) event that produced a SNR, now known as SN1006 or SNR G327.6+14.6. It is the brightest stellar event ever recorded and its historical record (Stephenson, 2010) is one of the reasons why this remnant was an interesting target for several observational campaigns. SN1006 is notable for being a relatively unobscured SNR (Katsuda et al., 2013) that is large in angular size due to its proximity to Earth (Winkler et al., 2003). All these points have made SN1006 a good object for studying SN1a events and has led to an impressive research history.

In particular, X-ray observations of the remnant have provided important information about the dynamics and energies of the supernova explosion and the surrounding interstellar medium. When a supernova explodes, it creates a rapidly expanding shell of ejected material that compresses and aggregates up the surrounding interstellar medium (ISM).

The collision between the expanding shell and the ISM creates a shock wave that propagates into the ISM and heats it up so that it emits thermal and non-thermal X-rays (Seward and Charles, 2010). In young SNRs, both thermal and non-thermal emission have a peak in the 0.5-10 keV energy range (Li et al., 2015), making current X-ray telescopes perfect for studying these objects. An important observation was made by Koyama et al. (1995), who detected synchrotron X-ray emission in the envelope of SN1006, supporting the theoretical expectation that the shock wave of SNRs accelerates particles to extremely high energies. This is believed to be a major production process of cosmic rays (CRs). Accordingly, SNRs are one important source of energy for the ISM via cosmic rays. This observation led to many subsequent spectral (Helder et al., 2012; Reynolds, 2008) and spatio-spectral analyses of SN1006 (Bamba et al., 2003; Winkler et al., 2014; Li et al., 2015) to study the spatially varying X-ray production processes in SN1006. In addition, supernovae are known to produce heavier elements from lighter ones during the explosion, which are ejected into the ISM and enable the formation of new stars and planetary systems, making them very important for the Galactic metabolism. Winkler et al. (2014) and Li et al. (2015) studied the spatial distribution of elements in the remnant. Long-term observations of SN1006 allowed Winkler et al. (2014) and Katsuda et al. (2013) to study its proper motion and thereby have offered insights into the dynamics of the explosion, the evolution of the remnant, and its interaction with the interstellar medium. Despite the extensive previous studies of SN1006 and other SNRs, there are still a number of aspects that are not well understood. Among them are the details of particle acceleration at shock fronts (Vink, 2011).

In recent years, there have been significant advances in X-ray astronomy aimed at studying such high-energy phenomena in the universe. These advances have been driven in large part by the development of new X-ray satellite missions such as Chandra, XMM-Newton and Suzaku, which have provided unprecedented spatial and spectral resolution. However, any technological advance in space-based astronomical instruments must be accompanied by advances in imaging methodology in order to exploit the full potential of these instruments. Here, we focus on the development of such an imaging method, capable of denoising, deconvolving, and decomposing the data, and apply it to Chandra observations of SN1006: the highest resolution data of this SNR to date. The aim is to obtain a more detailed view of the small-scale structures of the remnant and, thus, to allow a more detailed study of the open questions in the field of supernovae and their remnants as well as to challenge and benchmark the imaging method.

To obtain an accurate and meaningful reconstruction of the true flux from the given X-ray data, there are a number of challenges that need to be overcome. X-ray telescopes such as Chandra record the data from these high-energy phenomena as photon count events accompanied by information about the photon's arrival direction, time, and energy. In the present work, the events are categorized into spatial and energy bins, which yields independent Poisson statistics for each pixel. In particular, X-ray observations often have low count rates, which poses a challenge because of the resulting poor signal-to-noise ratio (S/N). Accordingly, a major task in X-ray imaging is the denoising of the corresponding data. In addition, there is an instrument-specific response to the observed X-ray flux,

which complicates the relation between the sky and data, particularly the exposure and the point spread function (PSF). A coherent representation and application of the response is a complex problem, as the instrumental properties of X-ray instruments tend to change with off-axis angle, energy, and time. Ultimately, one of the goals of X-ray imaging is to discriminate between noise, background, extended and point sources. So far, most imaging techniques are designed to extract either the point sources or the diffuse flux, but lack the ability to reconstruct both simultaneously. In the study of such extended sources as SN1006 in particular, this separation of components is essential when investigating the spectra and, thus, the emission properties of the remnant at each location.

The study presented here aims to address these challenges in X-ray imaging. In particular, we use information field theory (IFT) (Enßlin, 2019) as a versatile mathematical framework for reconstructing the signal from large and noisy data sets by combining information theory, statistical physics, and probability theory. Along with the numerical IFT algorithms implemented in the software package NIFTy (Arras et al., 2019), it provides an excellent tool for denoising, deconvolving and decomposing the image. This capability has already been demonstrated for Poisson data (Selig and Enßlin, 2015; Pompe et al., 2018; Scheel-Platz et al., 2023). The basis of IFT is Bayes theorem applied to the problem of reconstructing fields. In our case, the sky photon flux is regarded to be a field, which we subsequently refer to as the signal field. It is inferred given prior knowledge on its configuration and the measurement data, which is interpreted using a separate model for the measurement response. The instrument description, including its noise statistics determines the so called likelihood; in other words, the probability to observe specific data given a sky flux configuration. By combining the prior and the likelihood into the posterior distribution, we obtain not only an estimate of the actual sky photon flux as its posterior mean, but also an estimate of the uncertainty via the posterior variance.

During inference, the prior model guides the separation of the signal into different components, such as point-like and diffuse structures. Therefore, we need to carefully encode our knowledge of the different components into our prior model to give the inference the chance to discriminate their contributions to the observed photon counts. To this end, we have modeled the signal field as a superposition of different physical fluxes: the emission from point-like and extended sources. Assigning a different correlation structure to the diffuse emission from extended sources, which is assumed to be spatially correlated, and the point sources, which are assumed to be spatially uncorrelated, makes it possible to distinguish between these components. A spatio-spectral prior allows the reconstruction of the emissivity as a function of energy and spatial position. Further knowledge about the different spectra of the components improves their separation.

The instrumental description encoded in the likelihood drives the deconvolution of the data from the PSF, image denoising, and exposure correction. Specifically for Chandra, there are two different Advanced CCD Imaging Spectrometer (ACIS) X-ray imagers: ACIS-I and ACIS-S. The majority of the chips in ACIS-I and ACIS-S are front-illuminated. However, ACIS-S also contains two chips that are back-illuminated leading to a significant number of non-astronomical photon events in these regions. To account for the latter, we added a further model component of a non-astronomical, spatially varying but temporally

constant background that is present in regions of the back-illuminated chips. An additional challenge is the fact that Chandra’s field of view (FOV) is small compared with the extent of SN1006. It is therefore not possible to capture the entire remnant in a single ACIS-I or ACIS-S image. Instead, mosaicking is required (Winkler et al., 2014), which can be effectively implemented, even for varying instrument responses, by combining the corresponding likelihoods.

Overall, the spatio-spectral inference of the sky flux is associated with a significantly higher computational complexity than an inference that only considers the spatial direction. Therefore, we introduce a multi-step model, which considers two different priors: a purely spatial one and a spatio-spectral one. First, we performed a spatial reconstruction using the spatial prior. The result of this spatial reconstruction was mapped onto the entire spatio-spectral sky. The mapped sky with multiple energy bins added was then used as the initial guess for the subsequent spatio-spectral reconstruction. This allowed us to perform parts of the reconstruction and especially of the component separation in a smaller parameter space.

This multi-step model, which we call the transition model, and the reconstruction results on SN1006 are presented and discussed in this chapter. In section 3.2, we present current methods used in X-ray imaging and their application results on SN1006 data thus far. We also review state-of-the-art approaches to photon count data in the field of IFT. An introduction to the imaging of photon data with IFT is given in section 3.3. The explicit structure of the algorithm and in particular of the transition model is given in section 3.4. Section 3.5 focuses on the corresponding prior description and section 3.6 explains the instrument model and the Chandra observations of SN1006. In section 3.7, we present a reconstruction from synthetic data to validate the method, before finally presenting and discussing the reconstruction results on SN1006 in section 3.8. The conclusion and outlook for future research is given section 3.9.

3.2 Related work

This section is devoted to a review of previous studies and state-of-the-art developments in X-ray imaging. Previous investigations in high-energy astrophysics, with a focus on X-ray studies, are highlighted in three parts. First, we have a discussion of previous and current X-ray imaging techniques, followed by an explanation of the results of these techniques applied to SN1006 and an introduction to previous imaging techniques with IFT, which is the basis for the reconstruction presented here.

3.2.1 State-of-the-art X-ray imaging

The study of X-ray phenomena in the universe began in the 1960s and it is still a relatively new field of astrophysics due to the inability of ground-based telescopes to observe X-rays from astronomical sources. However, there have been many technical developments since then, discussed in greater detail in Seward and Charles (2010). Here, we focus on the

imaging techniques that have been developed and are in use, with a non-exclusive focus on the Chandra X-ray Observatory. A more comprehensive summary of recent developments in X-ray analysis for X-ray telescopes XMM-Newton, Suzaku, and Chandra can be found in Seward and Charles (2010).

Among others, Seward and Charles (2010) have offered insights into the steps and techniques in the widely used Chandra data processing pipeline¹. The corresponding methods and further data imaging and response tools have been implemented in the software tool Chandra Interactive Analysis of Observations (CIAO) (Fruscione et al., 2006), developed by the Chandra X-ray Center (CXC).

Overall, there are some standards for extended sources, such as SN1006, which have been applied in recent publications. One of these is the reduction of background from the data, which can obscure the signal from the source of interest. A disadvantage of this approach is that the subtraction of the background comes at the cost of eliminating real X-ray events. Another tool, applied in particular for extended sources, is mosaicking. This allows for the analysis of sources that have a greater extent than Chandra's FOV. In CIAO, mosaicking is implemented by transforming the raw count images, the effective area, and the background maps into a single coordinate system. Reconstructing an image from these mosaics has its difficulties, as there are often several PSFs and response matrix functions (RMFs) for one source. So far, this problem has been overcome by calculating and using the weighted average of the PSF and RMF for the data patches, as suggested by Broos et al. (2010).

One of the final steps, which depends on the object of interest, is source detection and extraction. The aim is to separate the X-ray source of interest from the background. For this purpose, three well-known algorithms have been implemented in CIAO: the sliding cell algorithm (Calderwood et al., 2001), wavelet detection algorithm (Freeman et al., 2008) and Voronoi tessellation and percolation algorithm (Ebeling and Wiedenmann, 1993). The sliding cell algorithm, previously used for Einstein and ROSAT, searches for sources by summing the counts in a square cell that slides over the image. For comparison, the counts in a cell assigned to the background are taken. From the ratio of the counts in the cell to the counts in the background, the cell might be assigned to a source. Wavelet detection, on the other hand, decomposes the signal into a series of wavelets. By analyzing the coefficients of the wavelets, patterns of different scales can be detected in the data. Finally, data cleaning and source extraction techniques differ for point sources and diffuse emission. This involves additional work as the pipeline needs to be run several times to fully extract point source and diffuse emission information.

There have been other approaches to source decomposition that fall into the category of blind source separation. In general, the goal of blind source separation is to automatically decompose observations into features maximizing their statistical separation. In Warren et al. (2005), a principal component analysis (PCA) approach was presented to determine the location of the contact discontinuity and the shock wave, and thus find evidence for cosmic ray acceleration in the SNR Tycho. In particular, sparse blind source separation

¹<https://cxc.harvard.edu/ciao/dictionary/sdp.html>

aims to compress the signal and extract its essence in this way. One application of sparse blind source separation to Chandra data was recently presented in Picquenot et al. (2019). Moreover, a generalized morphological component analysis (Bobin et al., 2016) was performed to X-ray data of Cassiopeia A by Bobin et al. (2020) to decompose the spectrum into its components such as thermal and non-thermal emission. Here, the generalized morphological component analysis models the source as a linear combination of a fixed number of morphological components and solves the according blind source separation problem, while putting sparsity constraints on the morphological components.

In addition, Bayesian and machine learning approaches have been applied for source separation, model comparison or point source characterization. Guglielmetti et al. (2009) analyzed Bayesian techniques for the joint estimation of sources and background, while Cruddace et al. (1988) implemented a maximum likelihood algorithm for the calculation of certain parameters of the detected sources, which has also been used for XMM-Newton. In Ellien et al. (2023), different components of the spectrum were modeled for Chandra data of five thin bands around Tycho and different one-, two-, and three-component models were analyzed via a Bayesian model comparison. Recently, a machine learning approach was published by Kumaran et al. (2023), with the aim of using it as an automated source classifier. The approach is based on supervised learning and allows point sources to be assigned to specific classes.

3.2.2 Previous studies of SN1006 in the X-ray range

The supernova remnant SN1006 has an exciting scientific record. As mentioned above, the remnant is of great scientific interest in the study of Type Ia supernovae and their remnants for many reasons: its proximity, low obscuration, and large size. In particular, X-ray observations of the remnant provide an opportunity to study its evolution. Accordingly, there have been intensive studies of SN1006 in this energy range, starting with observations by ROSAT (Willingale et al., 1996) and ASCA (Koyama et al., 1995). The ASCA data on SN1006 were analyzed by Koyama et al. (1995) and Dyer et al. (2003), which led to the confirmation of theoretical predictions that cosmic rays are accelerated at the shock fronts of the remnant. It was also discovered that there are several processes in the supernova remnant that are responsible for the X-ray emission. In fact, it was found that the northeast (NE) and southwest (SW) of SN1006 are dominated by non-thermal, synchrotron emission, while the northwest (NW) and southeast (SE) edges are less distinct and are attributed to thermal emission. Accordingly, Dyer et al. (2003) analyzed non-thermal and thermal models on the ASCA data.

The new technologies of the X-ray telescopes XMM-Newton and Chandra have led to an unprecedentedly high resolution of X-ray sources and, thus, to improved data for SN1006. Bamba et al. (2003) published the first spatio-spectral study of Chandra ACIS-S data from the NE shell of SN1006, followed by ACIS-I mosaic data from the analysis of Cassam-Chenai et al. (2008). Here, we want to highlight the publication of Winkler et al. (2014), as their reconstructed image ought to be the main point of comparison for ours. In Winkler et al. (2014), the standard Chandra pipeline (described in section 3.2.1) was used

and point sources were extracted using the wavelet detection algorithm. To use multiple data patches, the observations were merged using CIAO. A more recent study of SN1006 was carried out by Li et al. (2015) on XMM-Newton data using the SAS software. The data were preprocessed in a similar way to the Chandra pipeline and wavelet detection was used. However, point source detection was only possible at high energies because of the risk of misidentifying small-scale structures in the low energy regime as point sources.

3.2.3 Previous work on high energy count data with IFT

High-energy astronomical data, including X-ray and gamma-ray data, are recorded in photon counts. So far there have been no applications of IFT to Chandra X-ray data, but there have been studies on gamma-ray data and methodological research on the reconstruction and component separation of such count data. First, the algorithm D3PO by Selig and Enßlin (2015) based on IFT implemented the denoising, deconvolution, and decomposition of count data. Building on this, D4PO by Pumpe et al. (2018) allows D3PO to work on fields that have spectral and temporal coordinates in addition to spatial coordinates. Finally, Scheel-Platz et al. (2023) built a model of the gamma-ray sky and applied a variant of D4PO in a spatio-spectral setting.

In this work, we adapt a similar model as presented in Scheel-Platz et al. (2023) to describe the X-ray sky. As such, this is the first application of IFT imaging to X-ray data. Further, we introduce a method to fuse several data sets with different detector characteristics, pointing directions, and noise levels into a mosaic. We demonstrate how the imaging can be accelerated and improved by a multi-step model, which is presented in section 3.4.

3.3 Image reconstruction with IFT

In X-ray imaging, we are dealing with finite, incomplete, and noisy data. Here, we use IFT (Enßlin, 2019), an information theory for fields, to infer the X-ray sky as a continuous field from this finite data, d . In general, a physical field, $s : \Omega_z \rightarrow \Omega_s$, assigns a value to each point in the space Ω_z , which describes a continuous physical quantity such as temperature, pressure, intensity, and so on; in our case, this is the X-ray flux. Given the data, d , we obtain constraints on the field of interest, which we call the signal field. Since the data provide only a finite number of constraints on the signal field, there could have been an infinite number of signals that have produced the data, even if we completely neglect noise. For this reason, prior assumptions about the field are needed to sufficiently constrain the signal field, s . Given the likelihood, $\mathcal{P}(d|s)$, which describes the measurement, and a statistical description of the prior, $\mathcal{P}(s)$, the posterior probability of the signal given the data can be calculated via Bayes' theorem,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\mathcal{P}(s)}{\mathcal{P}(d)}. \quad (3.1)$$

Through IFT, we aim to inspect this posterior probability, as it allows us to draw posterior samples and thereby calculate any important posterior quantity such as the posterior mean,

$$\mu = \langle s \rangle_{(s|d)} \equiv \int \mathcal{D}s \, s \, \mathcal{P}(s|d), \quad (3.2)$$

where $\int \mathcal{D}s$ denotes the path integral over all possible field configurations and a measure of uncertainty via the covariance of the posterior probability,

$$D = \langle (\mu - s)(\mu - s)^\dagger \rangle_{(s|d)}. \quad (3.3)$$

The expectation value over the posterior probability is denoted by $\langle \rangle_{(s|d)}$ and \dagger gives the adjoint of the corresponding field. Therefore, the statistical treatment of the fields of interest in IFT creates an important advantage, as we may not only present a point estimate of the field, but also quantify its reliability at each position.

A more detailed description of the likelihood and prior model is given in section 3.4. Here, we describe image reconstruction with IFT given a general measurement equation. Accordingly, we consider a measurement as a function \mathcal{M} that maps a field from its continuous space to a discrete data space. This function is determined by the response, $\mathcal{R}(s)$, of the instrument and some statistical noise, n , in the measurement,

$$d = \mathcal{M}(\mathcal{R}(s), n). \quad (3.4)$$

Given this generic measurement equation we can calculate the likelihood by marginalizing over the measurement noise,

$$\mathcal{P}(d|s) = \int dn \, \mathcal{P}(d, n|s) = \int dn \, \mathcal{P}(d|n, s) \, \mathcal{P}(n|s), \quad (3.5)$$

$$= \int dn \, \delta(d - \mathcal{M}(\mathcal{R}(s), n)) \, \mathcal{P}(n|s), \quad (3.6)$$

$$= \mathcal{P}(\mathcal{M}^{-1}(\mathcal{R}(s), d)|s) \left| \frac{\partial \mathcal{M}(\mathcal{R}(s), n)}{\partial n} \right|^{-1}, \quad (3.7)$$

where \mathcal{M}^{-1} , is the inverse of the measurement function with respect to the second argument, n , and $|\partial \mathcal{M}/\partial n|$ is the functional determinant. When combining the likelihood with a prior distribution to obtain the posterior, the main difficulty lies in normalizing the posterior, namely, in computing the evidence: $\mathcal{P}(d) = \int \mathcal{D}s \, \mathcal{P}(d|s) \mathcal{P}(s)$. To circumvent the problem of analytically intractable normalization, we approximate the posterior via variational inference (VI), where a possibly complex posterior distribution $\mathcal{P}(s|d)$ is approximated by a simpler one, $\mathcal{Q}(s|d)$. Mathematically, the Kullback–Leibler divergence (KL) (Kullback and Leibler, 1951) is the measure that needs to be optimized to find the optimal approximation,

$$\mathcal{D}_{\text{KL}}(\mathcal{Q}(s|d) || \mathcal{P}(s|d)) = \int \mathcal{D}s \, \mathcal{Q}(s|d) \ln \left(\frac{\mathcal{Q}(s|d)}{\mathcal{P}(s|d)} \right). \quad (3.8)$$

Here, we use the VI version of the KL divergence; when minimized for the parameters of $Q(s|d)$, it ensures that in the approximation the least amount of spurious information is introduced. The expectation propagation (EP) version of the KL divergence, $\mathcal{D}_{\text{KL}}(\mathcal{P}(s|d)||Q(s|d))$, would be more conservative, as it would just ensure that a minimum of information is lost, but none have been introduced. However, EP requires integration over the intractable posterior $\mathcal{P}(s|d)$, while VI only requires integration over a conveniently chosen function $Q(s|d)$ (e.g., a Gaussian) and, therefore, this is feasible. As a consequence, uncertainty estimates obtained from the VI approximation are known to be a bit too optimistic, which should be kept in mind. However, those are nevertheless well informative about the structure of the uncertainties. For further details, we refer to Frank et al. (2021).

3.4 Algorithm overview of Bayesian inference of the X-ray sky

3.4.1 Structure of the reconstruction algorithm

The measure of interest in our reconstruction of SN1006 is the sky flux s as a function of space and energy. In other words, the signal field, s , lives on a space consisting of a 2D position space and a one-dimensional log-energy space, denoted by $z = (x, y) \in \Omega_z = \mathbb{R}^2 \times \mathbb{R}$, where $y = \log(\mathcal{E}/\mathcal{E}_0)$ and \mathcal{E}_0 is the reference energy. In order to guide the inference in the latent space and to reduce computational complexity, we introduce a multi-step model, which we call the transition model. The transition model divides the actual reconstruction into three parts, with three different inference problems, which are solved by VI. First, we aim to reconstruct the sky at a single energy level. Here, we perform a purely spatial reconstruction of the signal of interest. This part of the reconstruction is called the single frequency (SF) reconstruction. Its results are used to determine an informed starting position for the spatio-spectral reconstruction, subsequently called the multifrequency (MF) reconstruction. The standard reconstruction algorithm for the SF and MF model are further described in section 3.4.2. We model the mapping from the SF image to the MF image space as an inference problem, whose solution constitutes step two (introduced in section 3.4.3). In the third step, we solve the MF reconstruction using the starting point provided by step two. A similar model was previously used by Arras et al. (2022) to move from a spatial domain to a spatio-temporal domain.

Using the transition model, we can solve significant parts of the reconstruction problem, including the separation of point source and diffuse emission, in the SF setting, which has less model and computational complexity. Table 3.1 shows the number of hyperparameters for each model, SF and MF, and its sub-components, reflecting the model complexity. Table 3.2 presents the number of latent parameters in the model as a measure for the computational complexity. Figure B.2 and figs. B.3 to B.5 in the appendix show a quantitative comparison regarding computational complexity and reconstruction error for the transition model presented here versus a pure MF reconstruction. A schematic overview of the described reconstruction algorithm can be seen in the diagram in fig. 3.1.

The MF and SF prior models themselves are discussed in section 3.5.

3.4.2 Variational inference and generative models

As mentioned in section 3.3, we approximate the posterior given information on the prior model (as described in section 3.5) and the likelihood description (see section 3.6) via VI. Two approaches to VI of posteriors within the current NIFTy package are metric Gaussian variational inference (MGVI) (Knollmüller and Enßlin, 2020) and geometric variational inference (geoVI) (Frank et al., 2021). They are designed to approximate high-dimensional and complex posterior probability distributions via optimization of the cross entropy term of the KL in eq. (3.8). Both approaches perform the KL optimization in a coordinate space of the problem, in which the prior is a standard Gaussian. In particular, the signal field is described by a generative model $s = f(\xi)$ given a set of latent parameters ξ with a standard Gaussian prior $\mathcal{P}(\xi) = \mathcal{N}(\xi, \mathbb{1})$. The generative model encodes all prior knowledge on the corresponding field. To this end, the likelihood is formulated as a function of the latent parameters, $\mathcal{P}(d|\xi)$ and the posterior $\mathcal{P}(\xi|d)$ can be inferred via VI. In this work, geoVI is used, which optimizes the KL for the parameters of a non-linear coordinate transformation in which the posterior becomes an approximate standardized Gaussian. Thereby, geoVI allows for the representation of non-Gaussian signal posteriors. The detailed implementation can be found in (Frank et al., 2021). In any case, we need to define generative prior models for both, the SF and the MF model, given the corresponding latent parameters $s_M = f_M(\xi_M)$, $M \in \{\text{SF}, \text{MF}\}$. The detailed explanation of these models is part of the prior description in section 3.5. The according posterior approximations for each model M are denoted by \mathcal{Q}_M .

3.4.3 Transition model

The indirect encoding of fields in generative models complicates the transition from one model (e.g. the SF model) to another (e.g. the MF model) as the corresponding generative function is in generally not invertible; in other words, its inverse is not unique. Thus, our objective is to determine a mapping function that plausibly maps the parameters of the SF model, ξ_{SF} , to their corresponding MF parameters ξ_{MF} . As the transition model is intended to be flexible and adaptable to a range of initial and final models, we implement it as an inference problem. Given the posterior signal space mean $\mu_{\text{SF}} = \langle f_{\text{SF}}(\xi_{\text{SF}}) \rangle_{\mathcal{Q}_{\text{SF}}}$ and signal space variance $\sigma_{\text{SF}}^2 = \langle (f_{\text{SF}}(\xi_{\text{SF}}) - \mu_{\text{SF}})^2 \rangle_{\mathcal{Q}_{\text{SF}}}$ of the SF reconstruction, we infer the corresponding latent space parameters of the MF model, which we take as the starting point $\xi_{I,\text{MF}}$ for the MF reconstruction. The according virtual measurement equation is:

$$d_{\text{T}} = \mu_{\text{SF}} = \mathcal{R}_{\text{T}} f_{\text{MF}}(\xi_{I,\text{MF}}) + n, \quad n \sim \mathcal{N}(n, N), \quad (3.9)$$

where $N = \text{diag}(\sigma_{\text{SF}}^2)$. The transition response \mathcal{R}_{T} is a linear operator that can be chosen adaptively according to the problem under consideration. In the present analysis, \mathcal{R}_{T} is an operator that extracts the highest energy bin from the spatio-spectral field s_{MF} , generating

a 2D field, s_{SF} . The likelihood of the mapping inference problem is given by the linear measurement equation in eq. (3.9) and the likelihood derivation in section 3.3 is described by a Gaussian:

$$\mathcal{P}(d_{\text{T}}|\xi_{I,\text{MF}}, \sigma_{\text{SF}}^2) = \mathcal{N}(d_{\text{T}} - \mathcal{R}_{\text{T}}f_{\text{MF}}(\xi_{I,\text{MF}}), \text{diag}(\sigma_{\text{SF}}^2)). \quad (3.10)$$

The posterior for the initial latent parameters in the MF model $\mathcal{P}(\xi_{I,\text{MF}}|d_{\text{T}}, \sigma_{\text{SF}}^2)$ is approximated by $\mathcal{Q}_{\text{T}}(\xi_{I,\text{MF}}|d_{\text{T}}, \sigma_{\text{SF}}^2)$, with geoVI as described in section 3.4.2. We chose the posterior mean of the transition $\langle \xi_{I,\text{MF}} \rangle_{\mathcal{Q}_{\text{T}}}$ as the initial position for the subsequent MF reconstruction. This results in an overall algorithm that starts with a high-energy slice and uses this reconstruction as a starting point for the subsequent spatio-spectral reconstruction. The flow of reconstructions in this approach is illustrated in fig. 3.1. The decision to start with the high energy slice was deliberate, as this particular energy range has a more consistent effective area for Chandra. Since the transition result is used only as an initial guess, we assume a consideration of a diagonal transition noise covariance, N , to be sufficient. The possibly underestimated noise level is corrected in the subsequent spatio-spectral reconstruction steps.

3.5 Prior models for the X-ray sky

3.5.1 Prior composition

As described in section 3.4.1, we consider two different prior models, one for the SF reconstruction and one for the MF reconstruction. The fields in the SF reconstruction are defined in the spatial domain, $s_{\text{SF}} : \Omega_{\text{SF}} = \mathbb{R}^2 \rightarrow \mathbb{R}^+$, whereas the MF fields have an additional spectral dimension, $s_{\text{MF}} : \Omega_{\text{MF}} = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^+$. Regardless of the model, we assume that the X-ray sky consists of two possible sources: point sources and diffuse sources. Different prior models represent fluxes of different morphologies, each shaped by their physical production processes. The flux in diffuse structures should vary smoothly over position space. In other words, field values in the vicinity of a location are similar to that, which is best represented by the correlation structure of the field. In contrast, point sources are spatially uncorrelated and therefore best represented by spatially independent and sparsity enforcing priors. We discuss the specific use of either component in more detail below. In the following, the validity of the assumptions made for $s \in \{s_{\text{SF}}, s_{\text{MF}}\}$ is assumed to hold for both the SF and MF sky.

We represent the flux signal, s , as a superposition of point sources, s_{p} , and diffuse sources, s_{d} . In addition, we add a background component, s_{b} , which in our case accounts for the different backgrounds in front-illuminated (FI) and back-illuminated (BI) chips, which are further discussed in section 3.6.

Correspondingly, we denote the latent space sub-vectors, which parametrize these individual components, as ξ_{p} , ξ_{d} , and ξ_{b} , which altogether form the total latent space vector

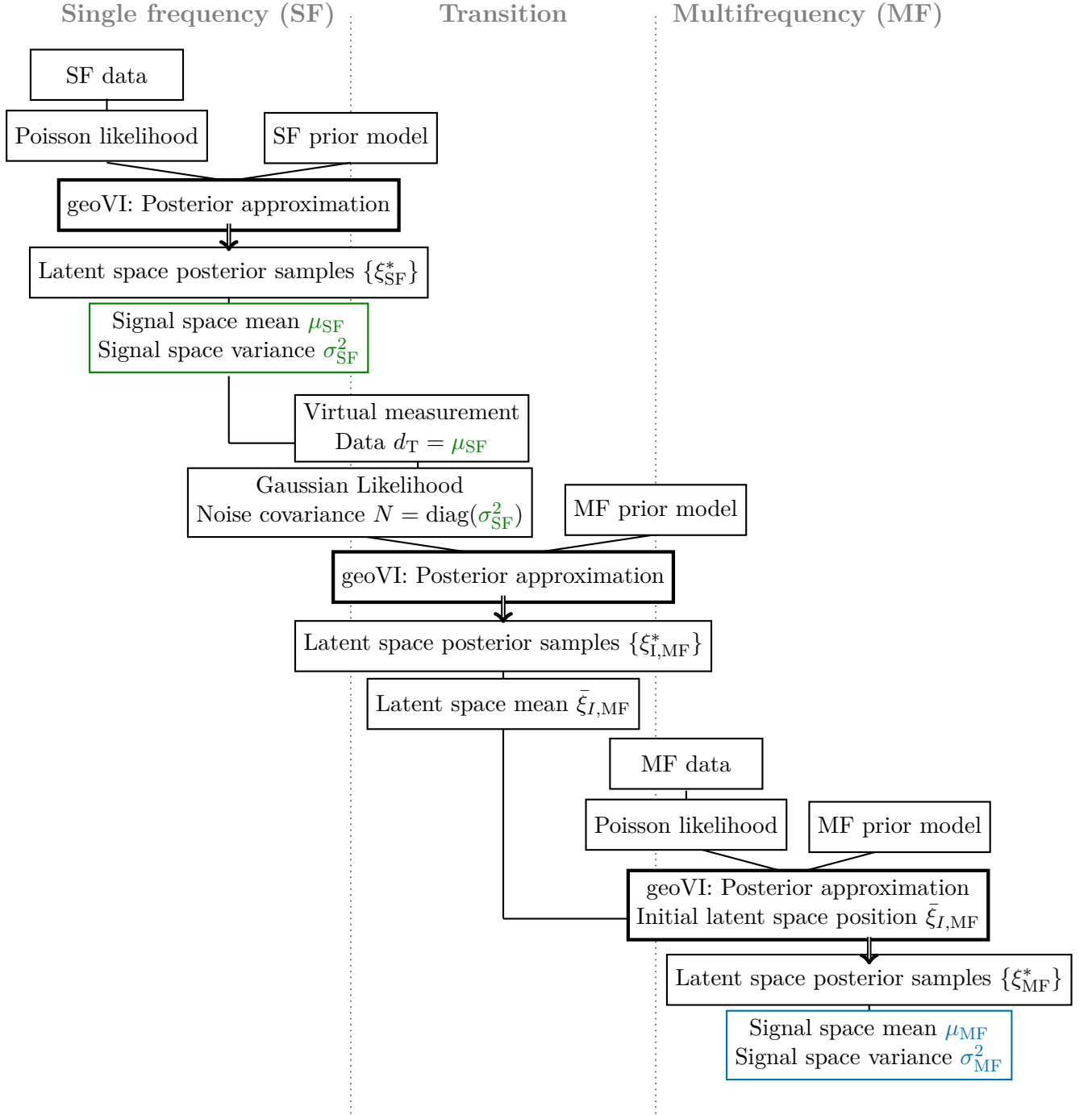


Figure 3.1: Structure of the transition model given the generative prior models for the SF reconstruction $s_{\text{SF}} = f_{\text{SF}}(\xi_{\text{SF}})$ and for the MF reconstruction $s_{\text{MF}} = f_{\text{MF}}(\xi_{\text{MF}})$, which transform the according latent parameters ξ_{SF} and ξ_{MF} from the latent space into the signal space. Here, $\xi_{\text{I, MF}}$ denotes the initial position of the MF reconstruction in latent space.

of the model, $\xi = (\xi_p, \xi_d, \xi_b)$. The according generative prior model is given by,

$$s = f(\xi) = \underbrace{f_p(\xi_p) + f_d(\xi_d)}_{\text{sky flux}} + \underbrace{\mathcal{R}' f_b(\xi_b)}_{\text{BI background}}. \quad (3.11)$$

Here, \mathcal{R}' denotes a mask, which assures that the additional background field is added in BI chip regions only. By expressing the transformation into the standardized coordinate system as a function s_i with $i \in d, p, b$, we obtain a generative model for each component as described in (Enßlin, 2022).

A set of prior samples can be seen in the first row of the synthetic data generation diagram in fig. B.1 in the appendix. Furthermore, the according prior samples and synthetic data allow us to choose the model hyper-parameters correctly. Here, we perform the search in two steps. First, we ensure mathematically via a coarse adjustment of the parameters that the order of magnitude in the counts of the data in fig. 3.3 is the same as the order of magnitude in the expected counts of the pixel-wise product of the exposure and the prior samples. Second, we look at the corresponding synthetic data, as shown in fig. B.1 and fine-tune hyper-parameters such that the components in the actual data and the synthetic data are morphologically similar, in order to improve the convergence of the algorithm.

3.5.2 Correlated components

Correlated components correspond to a flux that can vary over several orders of magnitude and exhibit spatial correlation. In this sense, the diffuse sky emission and the background are represented by correlated components. Their morphology is implemented by representing the signal for a correlated component as a log-normal process,

$$s = e^\tau \quad \text{with} \quad \mathcal{P}(\tau|T) = \mathcal{N}(\tau, T), \quad (3.12)$$

with an unknown covariance, T , describing the correlation structure of the correlated signal component. Since the correlation structure is not known a priori, we infer it concurrently by incorporating the correlated field model from Arras et al. (2022). Using the reparametrization trick introduced by Kingma et al. (2015), we describe the logarithmic sky flux as a generative process:

$$\tau = A\xi_\tau \quad \text{with} \quad T = AA^\dagger. \quad (3.13)$$

We went on to model the correlations in space and energy separately and assumed a priori statistical homogeneity and isotropy of the correlated logarithmic sky flux components in each of the subspaces $\Omega^{(k)}$, where $\Omega = \bigotimes_k \Omega^{(k)}$. Thus, according to the Wiener-Khinchin theorem, the corresponding covariances for the space $\Omega^{(k)}$, $T^{(k)}$ are diagonal and defined by the power spectrum $p_{T^{(k)}}$.

To learn the correlation structure of the correlated component, the power spectrum was modeled non-parametrically by representing the logarithmic power spectrum by an integrated Wiener process according to Arras et al. (2022) (The one-dimensional case

is further described in section 2.2.2). In particular, the mean and uncertainty of the parameters resulting from the chosen representation, such as the slope of the logarithmic power spectrum, its offset, and the fluctuations around the described power law, are learned from the data by modeling them as generative processes. This introduces further latent parameters to describe the generative model for the correlation structure. In the following, we refer to this prior model as the correlated field.

In the case of the SF model, we only considered the spatial correlations. Accordingly, the generative model for the spatially correlated components in the SF reconstruction is defined via $s_{\text{SF}} = f_{\text{SF}}(\xi_{\text{SF}}) = e^{\tau_{\text{SF}}(\xi_{\text{SF}})}$ with $\Omega = \Omega^{(k)} = \mathbb{R}^2$. In the MF model, we combined the power spectra for the independent spatial and spectral domain via a tensor product and define $s_{\text{MF}} = f_{\text{MF}}(\xi_{\text{MF}}) = e^{\tau_{\text{MF}}(\xi_{\text{MF}})}$. A further description of this generative model and its normalization can be found in Arras et al. (2022).

The diffuse and background components are represented by these spatially and spectrally correlated components. The number of derived hyper-parameters as well as latent parameters per component in each model is given in table 3.1 and table 3.2. For the correlation structure of the diffuse and background components, we made different prior assumptions to ensure an adequate separation of these components. In particular, we assumed that the spatial power spectrum of the diffuse sky structures has a slightly declining slope, allowing for small-scale structures in this component; meanwhile the spectrum of the background is assumed to be steep, allowing only for smooth background noise in the back-illuminated chips.

3.5.3 Point-like components

Point-like components appear local without any spatial correlation structure, due to their extreme distances. Consequently, we assume that the sky fluxes from point sources are spatially independent and, thus, their prior is factorized in a spatial direction:

$$\mathcal{P}(s_p) = \prod_x \mathcal{P}(s_p(x, y)).$$

In Selig and Enßlin (2015), different functional forms of possible point source luminosity priors were analyzed. Since the reconstruction of SN1006 requires a point source prior capable of modeling a few very bright point sources, we chose the inverse-gamma prior for the spatial direction according to Guglielmetti et al. (2009):

$$\mathcal{P}(s_p | \beta, \kappa) = \prod_x \frac{(\beta)^\kappa}{\Gamma(\kappa)} \left(\frac{1}{s_p^{(x)}} \right)^{\kappa+1} \exp\left(\frac{-\beta}{s_p^{(x)}} \right), \quad (3.14)$$

where κ is the shape parameter of the inverse-gamma distribution and β is the corresponding lower flux cutoff. The inverse-gamma prior behaves as a power law for fluxes much larger than the cutoff value, which matches the behavior observed for the luminosity functions of high-energy astrophysical point sources. Intuitively, it encodes the assumption

Model	s	s_d	s_p	s_b
SF	24	11	2	11
MF	53	19	15	19

Table 3.1: Number of hyper-parameters in each model per component

Model	s	s_d	s_p	s_b
SF	3.4×10^6	1.2×10^6	1.0×10^6	1.2×10^6
MF	1.4×10^7	4.4×10^6	5.0×10^6	4.4×10^6

Table 3.2: Number of latent parameters in each model per component

that with increasing luminosity, the set of point sources exceeding it becomes increasingly sparse. We model the inverse-gamma prior in standardized coordinates via inverse transform sampling leading to the generative model:

$$s_p = f_p(\xi_p), \quad (3.15)$$

where ξ_p is drawn from a standard Gaussian. Accordingly, f_p encodes the entire complexity of the inverse-gamma distribution and $f_p(\xi_p)$ is drawn according to eq. (3.14). In the SF model, eq. (3.15) describes the accurate generative model for the point sources.

For the MF model, we need to consider the spectral axis as well, by modeling the point source flux as spatially independent functions of the logarithmic energy y , according to Scheel-Platz et al. (2023). We assume that each point in the spatial subdomain has non-negligible correlations in the energy direction, as described by the correlated field component. In particular, we want to obtain a power-law dependence in the energy direction, defined by the spectral index α_p , and add fluctuations around it by a correlated field τ_p ,

$$(s_{\text{MF}})_p(x, y) = (s_{\text{SF}})_p(x) \frac{e^{\tau_p^{(x)}(y) + \alpha_p^{(x)} y}}{c}, \quad (3.16)$$

where c is the normalization. Here, not only the correlated field is described by a generative model, but also the spectral index $\alpha_p^{(x)}$ at every location x is learned. Thus, the additional energy axis introduces a number of new hyper- and latent parameters. The exact numbers of hyper- and latent parameters for the point sources in each model are given in table 3.1 and table 3.2.

3.6 Chandra instrument and data description

In Bayesian X-ray imaging, the prior model (section 3.5) is responsible for decomposing the components, whereas the denoising and deconvolution is controlled by the likelihood

model, which describes the measurement process. In general, an X-ray telescope provides photon counts that are statistically binned into pixels. This stochasticity is modeled by Poisson noise. The Poisson distribution gives the probability of the actual number of photon counts per bin, given the expected number of events, λ ,

$$\mathcal{P}(d|\lambda) = \prod_i \mathcal{P}(d_i|\lambda_i) = \prod_i \frac{1}{d_i!} \lambda_i^{d_i} e^{-\lambda_i}. \quad (3.17)$$

In the end, we want to know the photon flux at each point in position and energy space. To do this, we need to model the response function \mathcal{R} , which in a first step transforms the continuous flux field into a pixel-wise vector of expected photon counts λ_i , given a sky and BI background model. The response function includes all aspects of the instrument specific measurement, which are described in more detail below. Given the response, \mathcal{R} (see section 3.6.1), the number of expected counts at each pixel, λ_i , is calculated via $\lambda(z) = \mathcal{R}(s)(z)$. Because the Chandra FOV is small compared to the extent of SN1006, multiple observations were taken to cover the whole SNR. For each of several data patches, j , we get the data, d_j , and the response, \mathcal{R}_j , which need to be fused. Here, we introduce a mechanism that accounts for differences in the exposure and the PSF between the patches. By assuming that each patch is observed independently, we can write the log-likelihood of the full mosaic as the sum over individual patches (eq. (3.17)) for each data patch, d_j , and the corresponding expected counts, λ_j , calculated from the response, \mathcal{R}_j ,

$$\ln \mathcal{P}(d|\lambda) = \sum_j \ln \mathcal{P}(d_j|\lambda_j). \quad (3.18)$$

3.6.1 Chandra instrument response

We consider the data taken by the ACIS (Garmire et al., 2003), which is able to determine the energy of each incoming photon by using charge-coupled devices (CCDs). In particular, we consider the energy range 0.5 keV to 7.0 keV, which we bin in accordance with Winkler et al. (2014) into three energy bins (0.5-1.2 keV, 1.2-2.0 keV, and 2.0-7.0 keV). Chandra carries two different kinds of ACIS detectors, ACIS-I, used for imaging, and ACIS-S, used for imaging and spectral analysis. According to their application, ACIS-I and ACIS-S differ in the chips they are built of. In particular, ACIS-I is constructed out of FI chips only, which means that the incidental X-ray photons have to pass through the metal wiring until they reach the light-receiving surface. In contrast, ACIS-S also contains BI chips, where the CCD is flipped, such that the gate structure and channel stops do not face the X-ray-illuminated side. Accordingly, the BI chips are more sensitive to soft X-rays, so they are well suited for spectral analyses. However, they have a lower high-energy quantum efficiency and a worse resolution due to increased noise (Arnaud et al., 2011). The exact layout of ACIS-I and ACIS-S can be found in Chandra X-ray Center (2021).

We use version 4.14 of CIAO tool (Fruscione et al., 2006) designed by the CXC to extract information on the response ingredients such as the PSF and the exposure as well as on the event files itself for each patch. Here, we have made use of tools from the "data

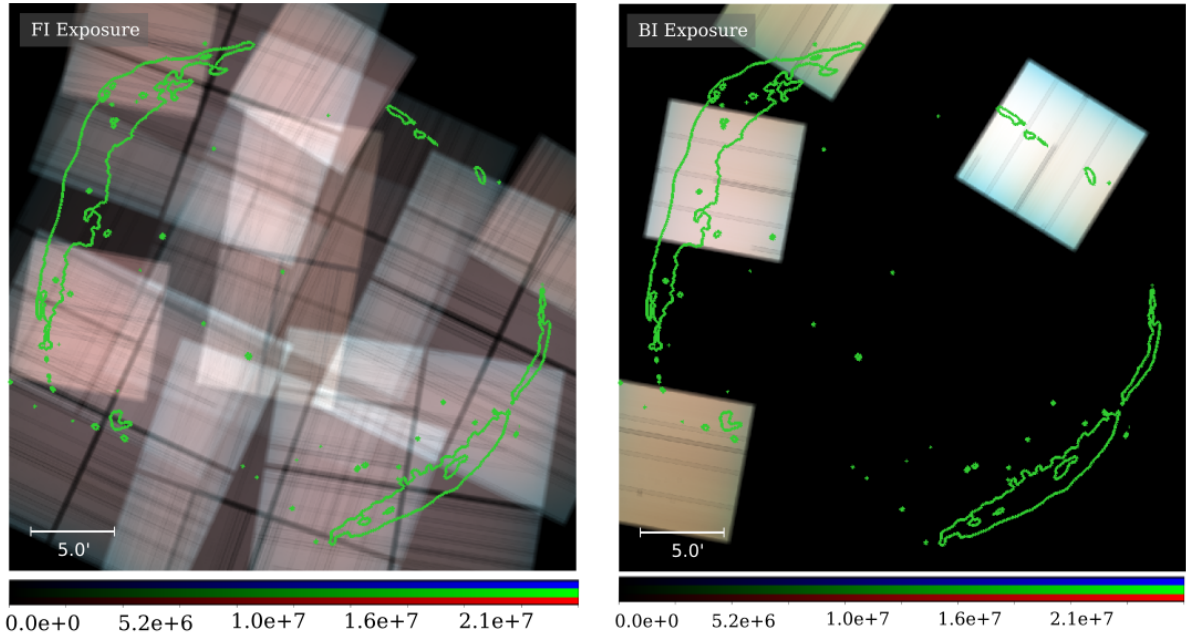


Figure 3.2: Visualization of the exposures used for the reconstruction in $[s \text{ cm}^2]$ (table 3.3): Top: Full exposure for all FI chips. Bottom: Full exposure for all BI chips.

manipulation" category for extracting and binning the data and from the "response tool" category to generate the ingredients of the instrument response. The exposure map is a key component in the process of converting raw X-ray data into scientifically useful data products, such as images and spectra. The exposure map combines information from the instrument map, which characterizes the instrument sensitivity such as the effective area and aspect solution (McDowell, 2006), which describes the spacecraft pointing and roll to create a map of the total observing time (or exposure) for each pixel in the field of view.

In Evans et al. (2010) the effective area as a function of energy is shown for the different chips. As mentioned above already, the FI chips are much less sensitive to low energy X-ray photons than the BI chips. On the other hand, the BI chips have more background flux. The exposure maps for the FI and BI chips can be seen in fig. 3.2. In order to account for the higher noise in the BI chips, we introduce an additional BI background field in section 3.5.

In general, for any of the considered chip types, it is evident that the chips are more sensitive to higher energies, leading us to the decision to take the highest energy bin (2.0-7.0 keV) as a starting point for the transition model. The PSF was simulated using the Model of AXAF Response to X-rays (MARX) (Davis et al., 2012). MARX is a software developed by the CXC (amongst others) to simulate the response (i.e., the PSF) of the Chandra X-ray Observatory, taking into account the telescope optics, pointing, and aspect of the telescope. We generated the response for each dataset and for each dataset, we used a homogeneous, spatially invariant PSF – but different PSFs for the different patches. The consequences of this approximation are addressed further in the quantitative discussion of

the results in section 3.8.1. For a further analysis of spatially variant PSFs, we refer to Eberle et al. (2023).

3.6.2 Chandra data of SN1006

The photon count data taken by the instrument is translated into an event table. Each event has information on time, energy, and position. Here, the data are binned into 1024×1024 spatial pixels and three energy bins. Moreover, the pointing direction of Chandra varies in time. Thus, the aspect correction is necessary, that is, taking into account the pointing direction of the telescope as a function of time. The data itself is an event list, which specifies the position in the chip coordinates and the arrival time of each photon. In McDowell (2001) the calculation of the sky coordinates of the photon given this event list is specified.

The latest Chandra observations of SN1006 according to Winkler et al. (2014) was chosen for the reconstruction presented here. Information on the data is given in table 3.3.

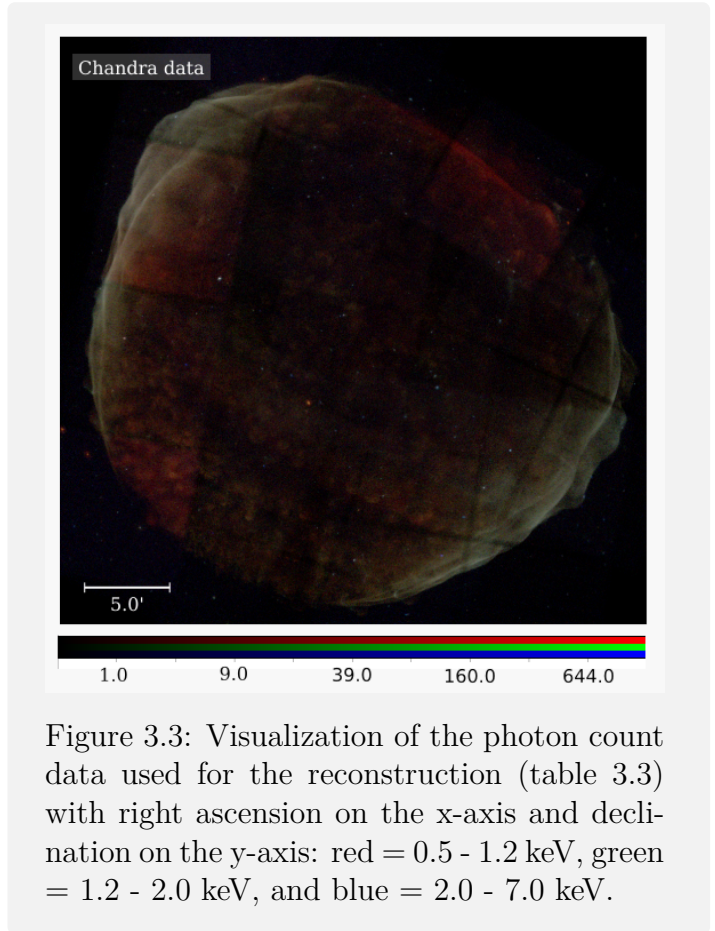


Figure 3.3: Visualization of the photon count data used for the reconstruction (table 3.3) with right ascension on the x-axis and declination on the y-axis: red = 0.5 - 1.2 keV, green = 1.2 - 2.0 keV, and blue = 2.0 - 7.0 keV.

The aim of the study of Winkler et al. (2014) was to measure the motion of the remnant and to get a more detailed view on its fine scale structures. Thus, the data observations were designed by Winkler et al. (2014) in order to match former observations, to be able to measure the expansion and using a longer exposure time in order to get a more detailed picture. The previous observations, which were taken as first-epoch images by Winkler et al. (2014), comprised a study of the non-thermal NE rim and the thermal NW rim with the ACIS-S array (Long et al., 2003; Katsuda et al., 2009) and the whole remnant with a mosaic of the ACIS-I observation (Cassam-Chenaï et al., 2008). Accordingly, the reconstruction deals with data from the BI and FI chips.

ObsID	Instrument	R.A.	Decl.	Date
9107	ACIS-S	15:03:51.5	-41:51:19	24.06.2008
13737	ACIS-S	15:02:15.9	-41:46:10	20.04.2012
13738	ACIS-I	15:01:41.8	-41:58:15	23.04.2012
14424	ACIS-I	15:01:41.8	-41:58:15	27.04.2012
13739	ACIS-I	15:02:12.6	-42:07:01	04.05.2012
13740	ACIS-I	15:02:40.7	-41:50:21	10.06.2012
13741	ACIS-I	15:03:48.0	-42:02:53	25.04.2012
13742	ACIS-I	15:03:01.8	-42:08:27	15.06.2012
13743	ACIS-I	15:03:01.8	-41:43:05	28.04.2012
14423	ACIS-I	15:02:50.9	-41:55:25	25.04.2012
14435	ACIS-I	15:03:42.5	-41:54:49	08.06.2012

Table 3.3: Information on the Chandra ACIS observations for the used data of SN1006 according to Winkler et al. (2014). Observations taken by the instrument ACIS-S are followed by the ACIS-I observations.

3.7 Validation of the algorithm using synthetic data

To demonstrate the performance of the developed algorithm, we perform the inference described above on a realistic but simulated dataset. Such a reconstruction based on synthetic data is useful not only for validating the reconstruction method, but also for determining certain parameters of the actual reconstruction, such as the sky flux detection limit. By constructing our model as a generative model, we were able to draw realistic sky samples that are similar to the region of the X-ray sky considered in this study. Given a sample of the sky, we were able to apply the response to it and mimic the Poisson noise. As a result, we obtained the synthetic data. The process of generating synthetic data is illustrated in fig. 3.4. For simplicity, we considered only three of the ACIS-I exposure patches for the synthetic reconstruction, rather than the whole set. The relevant synthetic data are shown in fig. 3.4, together with the actual simulated sky sample and the considered exposure map.

The resulting spatio-spectral reconstruction of the synthetic data in fig. 3.5 shows that the structures of the simulated sky are well captured. The denoising and response corrections are clearly visible when compared to the data shown in fig. 3.4. In particular, the right-hand side fig. 3.5 shows an enlarged version of the data and reconstruction to illustrate the denoising and deconvolution.

As mentioned before, a particular strength of the X-ray imaging method presented here is that we not only get an expectation of the signal, but also a corresponding standard deviation to this estimate. The corresponding pictures of the standard deviation for the individual energy bins can be seen in the appendix in fig. B.7. As expected, higher mean values exhibit greater variability in the flux, which in turn leads to a higher absolute uncertainty. However, given the standard deviation, σ_s , the reconstruction mean, μ_s , and

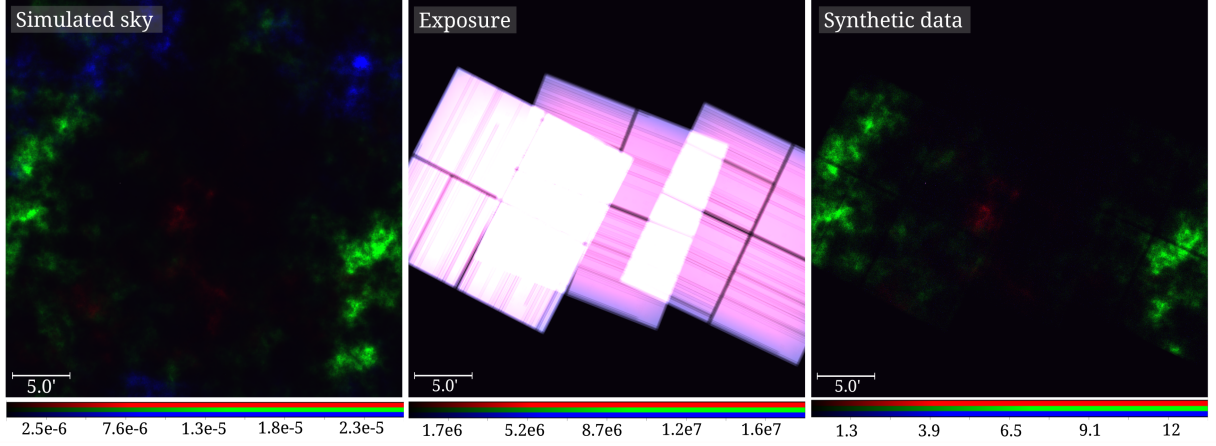


Figure 3.4: Generation of synthetic data: Left: Sky sample generated for the validation experiment. Center: Chandra exposure, modeled by combining three patches (14423, 14424, 14435) (see table 3.3). Right: Synthetic data corresponding to the sky sample, obtained by convolving the sky sample with the PSF and drawing a pixel-wise Poisson sample from the resulting detector flux prediction.

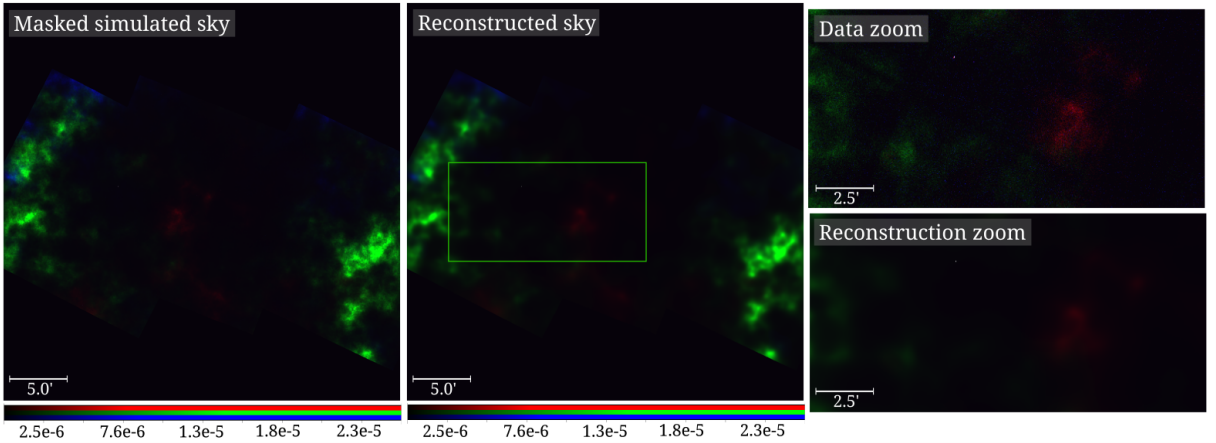


Figure 3.5: Reconstruction results on synthetic data: Left: Sky sample generated for this study masked by the extent of the exposure patches (14423, 14424, 14435) (see table 3.3). Center: Reconstruction result, i.e., the posterior mean, of the imaged sky masked by the extent of the same exposure patches. Right: Zoomed-in regions of the data on top and of the reconstructed image below. The shown cutout region is marked in the center image.

the fact that we know the signal ground truth, s_{gt} , itself from our generative model, we can calculate even more interesting validation measures, such as the UWRs per energy bin, i :

$$(\epsilon_{\text{UWR}})_i = \frac{(\mu_s)_i - (s_{\text{gt}})_i}{(\sigma_s)_i}. \quad (3.19)$$

In fig. B.6 of the appendix, we show the UWRs as well as the residuals,

$$\epsilon_r = (\mu_s - s_{\text{gt}}). \quad (3.20)$$

Areas with many counts show higher and more correlated residuals than those with low counts. Accordingly, the UWRs show that these pixels with high counts have higher uncertainty weighted residuals, due to a relatively small uncertainty. Overall, the simulated reconstruction demonstrates that the method developed is internally consistent.

Therefore, we used this synthetic reconstruction to set the threshold above which we can no longer distinguish noise from signal, which we refer to as the detection limit. The detection limit is used as a plotting lower limit in the actual reconstruction. Below this lower limit flux values are not shown in the image. The posterior approximation gives us the opportunity to draw posterior samples $s^* \leftarrow \mathcal{Q}(s|d)$, which we can use to calculate the sample averages. In order to determine the detection limit, we define the standardized error, ϵ_{rel} , between the ground truth, s_{gt} , and each of these samples, s^* , as a function of the ground truth flux,

$$\epsilon_{\text{rel}}(s_{\text{gt}}) = \left| \frac{s^* - s_{\text{gt}}}{s_{\text{gt}}} \right|. \quad (3.21)$$

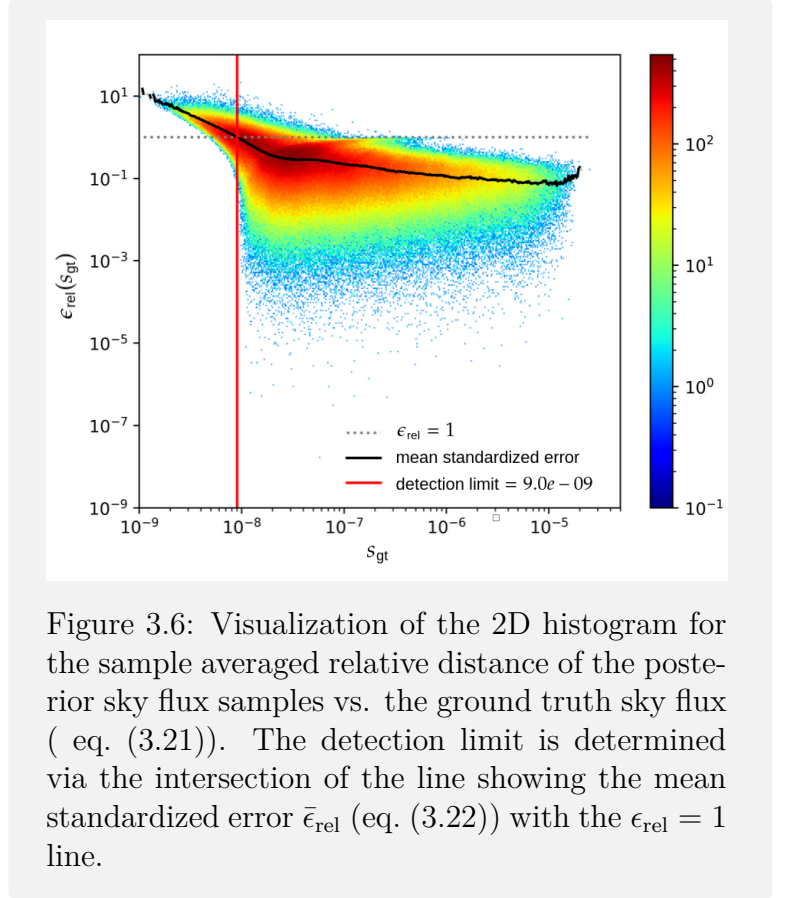


Figure 3.6: Visualization of the 2D histogram for the sample averaged relative distance of the posterior sky flux samples vs. the ground truth sky flux (eq. (3.21)). The detection limit is determined via the intersection of the line showing the mean standardized error $\bar{\epsilon}_{\text{rel}}$ (eq. (3.22)) with the $\epsilon_{\text{rel}} = 1$ line.

In fig. 3.6, the sample-averaged 2D histogram $\epsilon_{\text{rel}}(s_{\text{gt}})$ as a function of the ground truth flux, s_{gt} , is shown. For each value of $s_{\text{gt}, i}$ we calculate the mean standardized error of the histogram bins along the ϵ_{rel} -axis, $\bar{\epsilon}_{\text{rel}}(s_{\text{gt}, i})$, where $n(\epsilon_{\text{rel}, j}, s_{\text{gt}, i})$ is the number of counts for each bin (i, j) in the 2D histogram:

$$\bar{\epsilon}_{\text{rel}}(s_{\text{gt}, i}) = \frac{\sum_j \epsilon_{\text{rel}, j} n(\epsilon_{\text{rel}, j}, s_{\text{gt}, i})}{\sum_j n(\epsilon_{\text{rel}, j}, s_{\text{gt}, i})}. \quad (3.22)$$

Figure 3.6 reflects the expectation that the standardized error increases with smaller flux. To establish a detection threshold for low fluxes, we defined a limit beyond which we cannot confidently ascertain the presence of a signal in our observations. In this study, we

determined the detection threshold in such a way that the mean standardized error is less than 1. In other words, the detection threshold is set via the intersection point of the mean standardized error $\bar{\epsilon}_{\text{rel}}$ and the line $\epsilon_{\text{rel}} = 1$, leading to a detection limit of $9.0e^{-9}\text{s}^{-1}\text{cm}^{-2}$. We only performed this synthetic analysis on three of the data patches. Therefore, this threshold is conservative for the reconstruction with all patches.

3.8 Results and analysis of inference performance

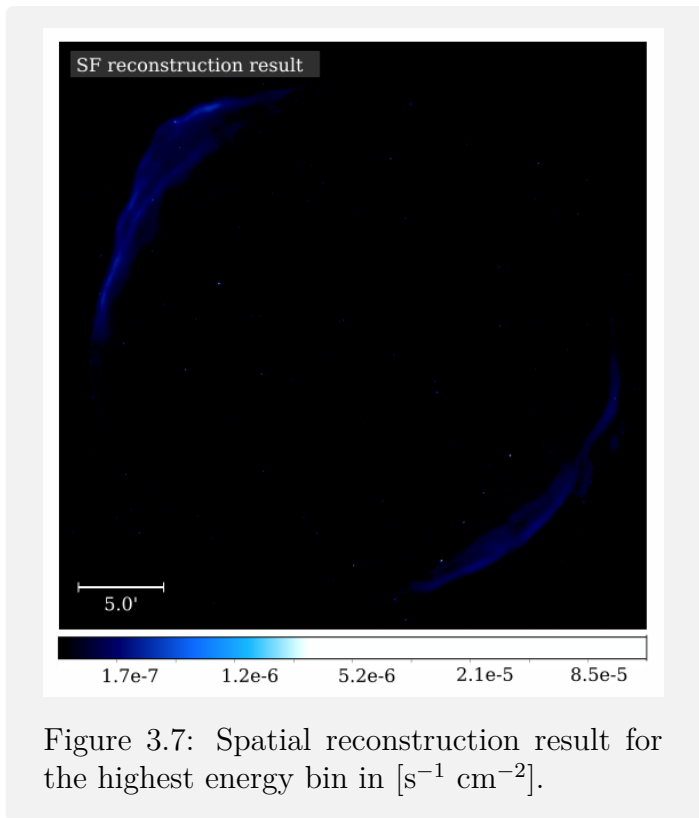


Figure 3.7: Spatial reconstruction result for the highest energy bin in $[\text{s}^{-1}\text{cm}^{-2}]$.

In this section, we discuss the results of the sky flux reconstruction. The additional background from the BI chips according to eq. (3.11) is removed in the reconstruction process. In fig. 3.7, we show the intermediate result of the SF reconstruction of the highest energy bin. This is taken as the initial condition of the subsequent MF reconstruction of SN1006, whose reconstruction results are shown in fig. 3.8 given the data shown in table 3.3 using Bayesian imaging and the transition model introduced in section 3.4.1. The reconstruction was visualized using the SAOImage DS9 imaging application (Joye and Mandel, 2003). The according results for each energy bin including the according color bars can be found in the appendix (fig. B.9).

As noted above, we are not only reconstructing the sky flux itself, but also its correlation structure in its correlated components. Accordingly, the posterior mean of the power spectrum in the spatial direction of the diffuse, extended sky component was also reconstructed and is shown for each energy bin in the appendix in fig. B.8.

3.8.1 Quantitative discussion

Figure 3.8 displays the final results of applying our reconstruction algorithm to SN1006: the reconstructed sky and its separated components. The overall separation of diffuse emission and point sources was successful. The point sources are clearly identified and the PSF deconvolution is particularly evident for the point sources. The diffuse structures are almost free of point source contributions.

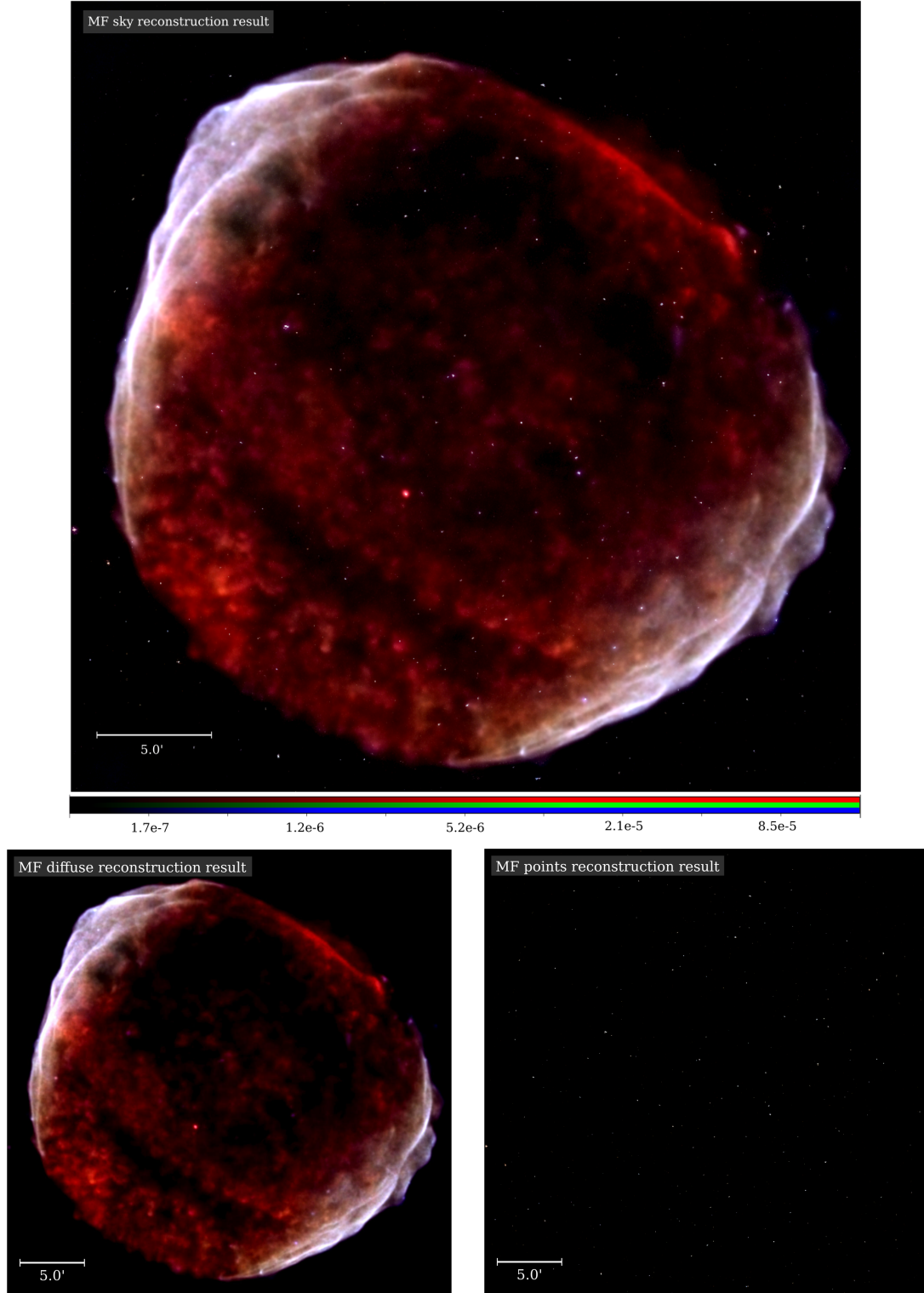


Figure 3.8: Reconstruction results for the flux in $[s^{-1} \text{ cm}^{-2}]$ (red = 0.5-1.2 keV, green = 1.2-2.0 keV, and blue = 2.0-7.0 keV. The corresponding color bars for each of these energy bins can be found in fig. B.9): Top: Full sky reconstruction mean. Bottom left: Reconstruction mean for diffuse emission. Bottom right: Reconstruction mean for point sources.

The effect of component separation can be seen more clearly in fig. 3.9, which shows a zoom-in on the NE quarter of the remnant and its components. It can also be seen that some of the additional background noise from the BI chip is partially absorbed into the background model.

One soft X-ray point source in the center of the remnant was not well separated from the diffuse emission. We believe this was caused by PSF mismodeling in the outer pixels of the detectors, where the source is located in all observations considered, due to the assumption of invariant PSFs within each observation. Figure 3.10 shows the exposure maps of the observations that covered this source and the position of the point source within these exposure maps. The pointing of ACIS-I and ACIS-S described in Chandra X-ray Center (2021) suggests large deviations of the actual PSF from the PSF model used in the positions of the mismodeled point source in detector coordinates. Dealing with position dependent PSFs will be addressed in a future publication.

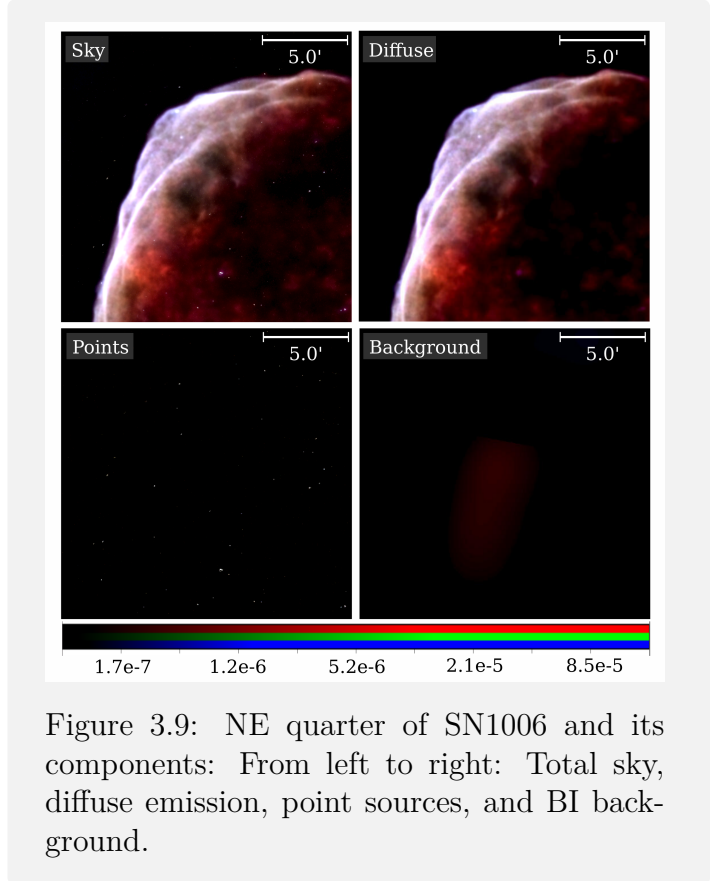


Figure 3.9: NE quarter of SN1006 and its components: From left to right: Total sky, diffuse emission, point sources, and BI background.

Figure 3.11 shows the reconstruction of diffuse emissions from the remnant in detail. In order to study the remnant effectively, it is crucial to get a detailed view of it. To improve the clarity of our results, we present four close-up images of the remnant, highlighting its small-scale structures. The analysis shows that the shell is denoised both in the NW region (where we expect thermal emission) and in the SW region (where we expect non-thermal emission). We can also see that the denoising has improved the resolution of the small-scale structures in the inner X-ray emission of the remnant with respect to the data and also in comparison to the previous study of Winkler et al. (2014).

Due to the statistical approach presented in this study, we obtained an estimate of the sky flux via the mean of the posterior probability, but also an uncertainty estimate via its standard deviation. The corresponding standard deviations for each energy bin are shown in the appendix in fig. B.9. The top row of the figure shows the different energy bins of the posterior mean for better comparison. It can be seen that, as expected, the standard deviation is higher for regions of higher flux.

Thanks to the probabilistic approach we adopted, we gain the ability to draw posterior samples from the inferred distribution. Such posterior samples, s^* , have allowed us to compute the posterior mean, the standard deviation or any other statistical quantity of interest. Correspondingly, we can calculate the absolute NWR, $(\epsilon_{\text{NWR}})_j$, for each data patch, j , as another interesting outcome of our results:

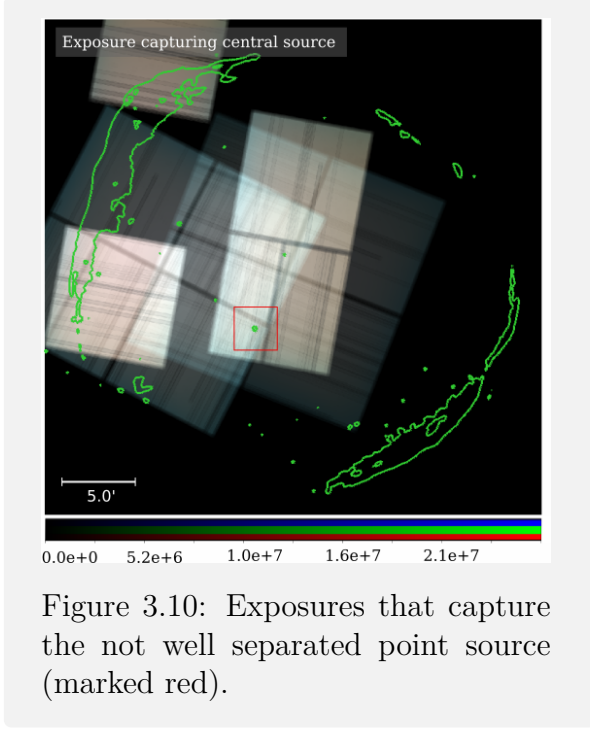


Figure 3.10: Exposures that capture the not well separated point source (marked red).

$$(\epsilon_{\text{NWR}})_j = \left\langle \frac{|d_j - \lambda_j(s)|}{\sqrt{\lambda_j(s)}} \right\rangle_{\mathcal{Q}(s|d)} \quad (3.23)$$

$$\approx \frac{1}{N} \sum_{i=1}^N \frac{|d_j - \lambda_j(s_i^*)|}{\sqrt{\lambda_j(s_i^*)}}. \quad (3.24)$$

Here, λ_j describes the reconstructed expected number of counts for each pixel in the data patch, j . The absolute NWRs provide a way to quantify the difference between a measured data point and its reconstruction, and help to distinguish between true deviations of the data from the reconstruction and deviations that are simply due to Poisson noise.

The plots of the absolute NWRs are shown in the appendix in table B.1 for each energy bin and data patch. These plots can be used as a sanity check on the correctness of the reconstruction presented here, as they allow us to point out systematically unmodeled effects in the likelihood and the prior. We can see that the NWRs are close to one in most regions, which implies a well-fitting model and reconstruction. In particular, in regions around point sources or at strong edges, we find higher NWRs, which we attribute to the well-functioning deconvolution in these regions, leading to deviations of the reconstructed signal from the data.

3.8.2 Analysis and comparison with previous studies

As mentioned above, SN1006 has been studied extensively using a variety of instruments. In particular, several studies using X-ray telescopes have produced images of the SNR. These studies have provided important insights into its structure and evolution, thus advancing our understanding of supernova explosions and their aftermath. Corresponding imaging approaches to SN1006 can be found in Winkler et al. (2003), Li et al. (2015), Bamba et al. (2003) and also in Fruscione et al. (2006), which demonstrated the fidelity of the Chandra

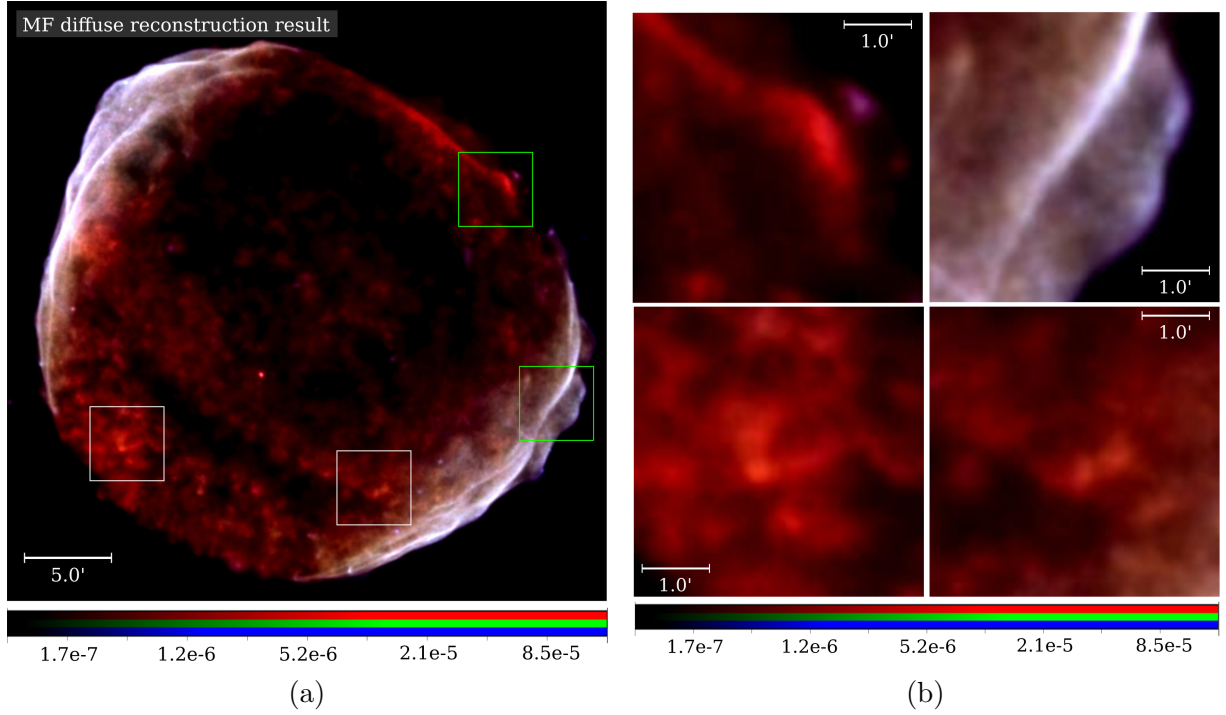


Figure 3.11: Zoom-in on the reconstruction results in the diffuse component: (a) the whole diffuse reconstruction and the locations of the zoomed-in areas; (b) the top panels represent the green areas marked in the remnant and show zoom-ins on the denoised shell of the remnant. The lower panels are represented by the white boxes in (a) and show structures in the inner soft X-ray emission of the remnant.

data processing pipeline for SN1006. In particular, in this study we have focused on the data and energy regions used by Winkler et al. (2014) and compare our reconstructions with their results. In Winkler et al. (2014), a comparison was made between the X-ray image and a $H\alpha$ image of SN1006 from CTIO. The comparison shows that there are several thin arcs of Balmer emission in the southern part of the remnant, which lie just in front of small-scale structures in the X-ray emission. In fig. 3.11, we show these central parts of the remnant, which are dominated by soft X-rays. We show the enlarged cut-outs of these regions for the extracted remnant. Compared to previous studies, small-scale structures have an improved resolution due to the denoising and deconvolution and are well disentangled from any point sources in the background. Accordingly, the presented reconstruction provides a more detailed view of the inner part of the remnant, enabling a more accurate study of its small-scale structures.

As noted in Li et al. (2015), the different energy bands show spatial variations in the remnant SN1006 with respect to each other. Figure 3.8 shows these differing spatial variations of intensity with different X-ray energy bands. In particular, in fig. 3.9 and fig. 3.11, parts of the hard X-ray lobe (the non-thermal regions in the SW and NE of the remnant) are well resolved and denoised, without any point source contribution. Soft X-

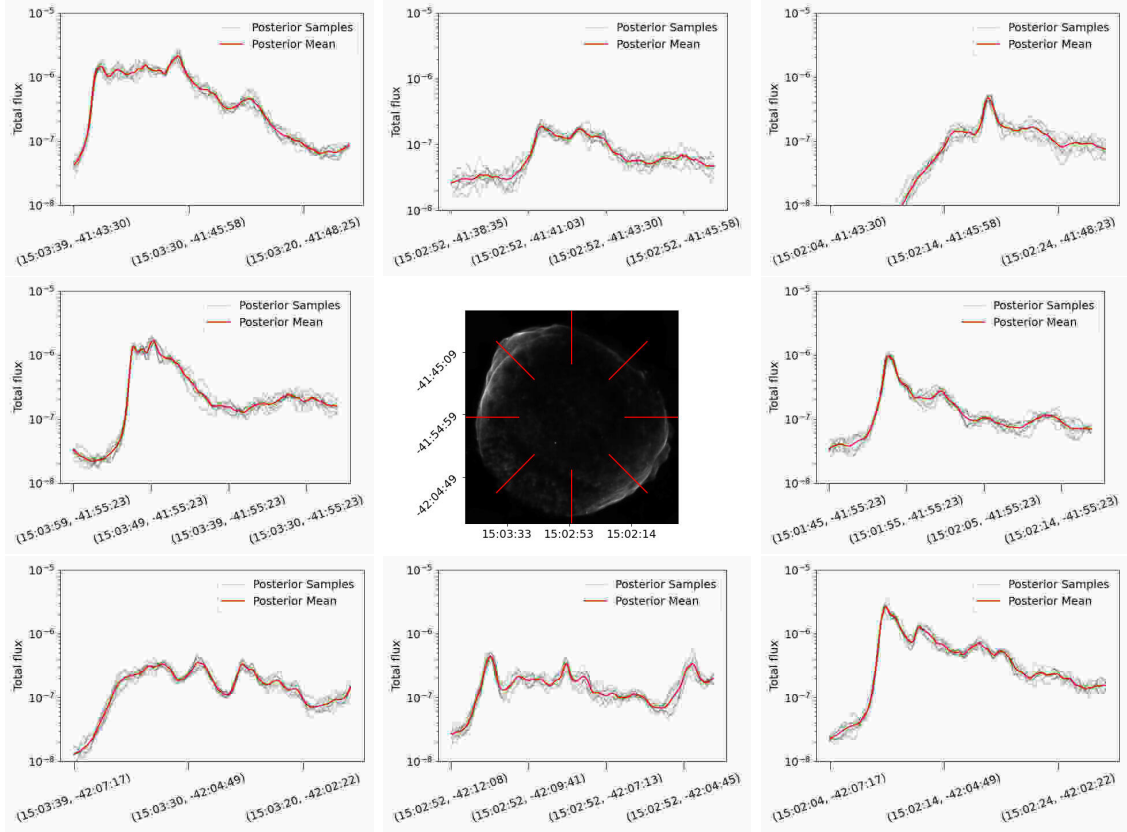


Figure 3.12: Flux intensity profiles in $[s^{-1} \text{ cm}^{-2}]$: The center image shows the location of the lines along which we present the intensity profile in pixel coordinates. The corresponding intensity profiles are plotted next to the line. The posterior mean of the reconstructed flux is plotted in red and the corresponding posterior samples are plotted in grey. The profiles are shown from left to right from the outside to the inside of the SNR.

rays in the NW shell are shown in fig. 3.11, which shows the shell of the thermal emission. Koyama et al. (1995) presented the first observational evidence that supernova shocks produce cosmic rays. However, the details of the acceleration mechanism of the particles is still an open question. Therefore, the study of the shock fronts is important to gain further insight into the acceleration mechanism and the dynamics of the shock front. The separation of the diffuse emission from the remnant allows us to visualize long intensity profiles along the remnant. Figure 3.12 shows such radial intensity profiles of the SNR in eight equidistantly space orientations. These denoised and deconvolved profile lines can be very useful to search for halos in front of the non-thermal regions and to get insights into the magnetic field strength according to Helder et al. (2012). Here, the profile lines show strong and sharp X-ray flux increases by up to two orders of magnitude at the shock front in the non-thermal regions in the SW and NE of SN1006. Notwithstanding, we defer the analysis of the structure of the remnant based on our reconstructions to future works.

3.9 Conclusion

In summary, we present a technique for obtaining an estimate of the true sky photon flux from Chandra X-ray event data. By using the sky flux as a generative process, this method allows us to infer not only the flux itself, but also its correlation structure in its extended components. Based on IFT and the application of Bayes' theorem, this method approximates the posterior probability of a signal given the data via geoVI in a problem-adapted latent space. This allows us to draw posterior samples in order to compute the expected sky flux, posterior uncertainty, and further validation and diagnostic measures.

Modeling the different correlation structures of point sources, diffuse emission, and background in the BI chips, we were able to separate point-like, extended sources and the additional noise in the BI chips from the sought-after signal. Compared to previous separation and source extraction techniques, which are usually specified to extract either point sources or extended sources, the inference based on IFT accounts for both components jointly. In particular, we built a spatio-spectral model for the sky flux based on the D4PO algorithm implemented by Scheel-Platz et al. (2023) and used it for the spatio-spectral reconstruction of the X-ray sky. Since the spatio-spectral reconstruction is computationally expensive for a large number of pixels, we introduced an accelerated inference model, called the transition model. In the transition model, we first performed a spatial reconstruction in a single energy band, which has almost one order of magnitude less degrees of freedom as the spatio-spectral reconstruction, making it computationally less complex. The result of the spatial reconstruction, which already contains a lot of information about the sky flux in an energy bin and about the component separation, was used as an initial condition for the spatio-spectral reconstruction.

A benefit of the here presented analysis is the ability to build mosaics of different observations via the sum of logarithmic likelihoods. Each likelihood has its own description of the instrument response. This approach solves the problem of modelling different PSFs for the same source in different data patches.

We applied the spatio-spectral reconstruction to the latest Chandra observations of SN1006, presented by Winkler et al. (2014). The resulting image is a denoised, deconvolved and decomposed image, which provides a detailed view of the small-scale structures of SN1006. We reconstructed a separate image of the point sources present in the considered datasets, which can be compared with point source catalogs and, more importantly, which allows us to study the X-ray emission of SN1006 without point source contributions. The different energy ranges in NE and SW – dominated by synchrotron emission – and the rest of the remnant – dominated by thermal emission – are clearly visible. The intensity profiles at the shell of the remnant are denoised and not visibly affected by point source contributions. We also show other diagnostics such as a simulated data reconstruction, uncertainties, and noise-weighted residuals as a check for systematic errors.

Taking this work as a starting point for spatio-spectral Bayesian imaging of X-ray data, we have pointed out the need for further methodological improvements. One is based on the use of a spatially varying PSF. Alexander et al. (2003) already showed the actual spatial variability of the PSF in the Chandra image of the Deep Field North. We expect that the

separation of the central point source in SN1006 to improve by the implementation of a spatially varying PSF. However, this is not a trivial task, as an invariant PSF can be applied via multiplication in Fourier space, whereas a spatially varying PSF cannot. Methods are currently being developed to solve this problem in the language of IFT, including a neural network approach recently presented by Eberle et al. (2023). In addition, a line model capable of modelling lines in the thermal emission will help to further resolve the energy direction. An interesting option here, already mentioned in Seward and Charles (2010), would be to define different models for synchrotron emission and bremsstrahlung, with the goal of eventually decomposing the diffuse emission of the remnant into its thermal and non-thermal components. In general, we aim to further improve the reconstruction speed and reduce its computational cost to enable studies of more data sets, larger regions, and with higher resolutions in the spatial and spectral dimensions. In addition, we want to further optimize our hyper-parameter search to enable a faster convergence of the algorithm.

3.10 Acknowledgements

M. Westerkamp, V. Eberle and M. Guardiani acknowledge support for this research through the project Universal Bayesian Imaging Kit (UBIK, Förderkennzeichen 50002103) funded by the Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR). P. Frank acknowledges funding through the German Federal Ministry of Education and Research for the project ErUM-IFT: Informationsfeldtheorie für Experimente an Großforschungsanlagen (Förderkennzeichen: 05D23EO1). J. Stadler and J. Knollmüller acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311.

Chapter 4

Bayesian multi-wavelength imaging of the eROSITA LMC SN1987A

*The following chapter is a manuscript that has been submitted to **Astronomy & Astrophysics** with Vincent Eberle, Matteo Guardiani and me as the shared first authors (Eberle et al., 2024a). In close collaboration Vincent Eberle, Matteo Guardiani, Philipp Frank, and I implemented the first inference pipeline in NIFTy for a spatial reconstruction of the eROSITA sky. This was facilitated by the eROSITA data loading pipeline implemented by Matteo Guardiani. The theoretical calculations and implementation of the spatially variant PSF used were carried out by Vincent Eberle. The entire code was rewritten by Vincent Eberle, Matteo Guardiani and myself to be JAX-compliant using NIFTy.re. To this end, Vincent Eberle and I implemented a JAX-compatible spatio-spectral prior model. Matteo Guardiani and I built an extensive diagnostic pipeline for the results. Other contributions of mine were the implementation of extended sources in the prior model, the tuning of the hyperparameters for the final reconstruction, and the corresponding validation of the algorithm in a simulated data reconstruction. The scientific exchange with Michael Freyberg and Mara Salvato about the eROSITA data and instrument has been very beneficial to the presented work. The whole work would not have been possible without many fruitful discussions with Torsten Ensslin. The first draft of the manuscript was written jointly by Vincent Eberle, Matteo Guardiani, and myself. The individual contributions to the first draft are marked within the chapter. The final manuscript was refined and reviewed by the authors.*

For consistency within this thesis, some of the parameters have been renamed and some of the figures have been adapted according to the layout.

The EDR and eRASS1 data have already revealed a remarkable number of undiscovered X-ray sources. Using Bayesian inference and generative modeling techniques for X-ray imaging, we aim to increase the sensitivity and scientific value of these observations by denoising, deconvolving, and decomposing the X-ray sky. Leveraging information field theory, we can exploit the spatial and spectral correlation structures of the different physical components of the sky with non-parametric priors to enhance the image reconstruction. By incorporating instrumental effects into the forward model, we develop a comprehensive Bayesian imaging algorithm for eROSITA pointing observations. Finally, we apply the developed algorithm to EDR data of the LMC SN1987A, fusing data sets from observations made by five different telescope modules. The final result is a denoised, deconvolved, and decomposed view of the LMC, which enables the analysis of its fine-scale structures, the creation of point source catalogs of this region, and enhanced calibration for future work.

4.1 Introduction

First draft written by Vincent Eberle.

The eROSITA X-ray Telescope (Predehl et al., 2021) on Spectrum-Roentgen-Gamma (SRG) (Sunyaev et al., 2021) was launched on July 13th, 2019 from the Baikonour Cosmodrome. Since the full-sky survey of ROSAT (Truemper, 1982) in 1990, eROSITA is the first X-ray observatory to perform a full-sky survey with higher resolution and a larger effective area. After a calibration and phase verification (CalPV) phase of pointed and field-scan observations, the main phase of the mission is devoted to multiple all-sky surveys of the celestial sphere, each lasting about 6 months. The amount of data collected by the X-ray observatory in its about 4.3 completed all-sky surveys already has a huge scientific impact. In order to make use of scientific data, nuisance effects of the instrument need to be understood and removed whenever possible. Amongst others, Poissonian shot noise and the point spread function (PSF) of the optical system cause problems to source detection algorithms. Unfortunately some of these effects are not analytically invertible and thus leave us with an ill-posed problem at hand. Information field theory (IFT) (Enßlin et al., 2009) provides us with a theoretical framework that allows to address these kinds of problems. The use of prior knowledge and generative modeling enables us to remove instrumental effects, decompose the sky into physical components, potentially remove the high-energy particle background and leave us with a statistical approximation of the posterior distribution. This permits us to gain knowledge about any posterior measure of interest, such as the mean and the uncertainty of the measured physical quantities.

4.1.1 Related work

First draft written by Margret Westerkamp.

X-ray astronomy has developed rapidly since its beginnings in the 1960s, driven by major X-ray missions such as Einstein and ROSAT. This rapid progress has been fueled not

only by advancements in instrumentation, with ever-improving telescopes such as Chandra (Garmire et al., 2003), XMM-Newton (Schartel and Dahlem, 2000), and more recently eROSITA, but also by simultaneous developments in imaging techniques. These advancements have steadily increased the amount of information extracted from observations and enabled researchers to address various data-analysis challenges. For instance, tasks such as source detection and the coherent fusion of overlapping datasets – some of the most difficult tasks in astrophysical imaging – along with the denoising and deconvolution of X-ray data affected by Poisson noise, have become more manageable due to these innovations.

In chapter 3, an overview of general source detection algorithms is given, such as the sliding cell algorithm (Calderwood et al., 2001), the wavelet detection algorithm (Freeman et al., 2008) and the Voronoi tessellation and percolation algorithm (Ebeling and Wiedenmann, 1993), as well as an overview on data fusion techniques currently used and implemented for Chandra data. A summary of the data processing and imaging pipelines for the Chandra and XMM-Newton X-ray telescopes is available at Seward and Charles (2010). This section provides an overview of the state of the art in X-ray imaging techniques specifically for eROSITA. For eROSITA data analysis, there is the eROSITA Science Analysis Software System (**eSASS**) (Brunner et al., 2018; Merloni et al., 2024), which includes all the functionalities of the standard eROSITA processing pipeline, such as event processing, event file and image generation, background and point source detection, and source-specific output such as light curve and spectrum generation. In Brunner et al. (2022) the standard eROSITA source detection pipeline using **eSASS** for eROSITA Final Equatorial-Depth Survey (eFEDS) is elaborated step by step. First, the standard source detection requires a preliminary source list containing all possible source candidates, which is generated using the sliding cell algorithm. Based on the preliminary source catalog the X-ray data is compared to an PSF model using maximum likelihood fitting. Finally, as noted in Merloni et al. (2024), the point sources can be excluded by a circular region centered on the point sources with a certain radius to obtain a point source subtracted images. All the necessary functionalities, including the sliding cell algorithm, are implemented in the corresponding **eSASS** package. To test the completeness and accuracy of this source detection pipeline, Liu et al. (2022) simulated eFEDS data by simulating the instrument response, given a specific catalog and background, and applied the source detection algorithm to it.

Recently, Merloni et al. (2024) published a catalogue of point sources and extended sources in the western Galactic hemisphere using the first of the all-sky scans of eROSITA All-Sky Survey (eRASS1). In this study we focus on the imaging of the LMC using EDR data from the **CalPV** phase. As the nearest star-forming galaxy, the LMC has already been observed and analyzed in its various parts across the entire electromagnetic spectrum, as noted in Zangrandi et al. (2024); Zanardo et al. (2013). Among other things, the numerous supernova remnants (SNRs) present in it are of interest, as studied in Zangrandi et al. (2024) on data from eRASS:4, including all data from the eROSITA all-sky surveys eRASS1-4. To enhance the edges of the shocked gas in the SNRs, they used the Gaussian gradient magnitude (GGM) filter (Sanders et al., 2016), resulting in 78 SNRs and 45 candidates in the LMC. The most famous supernova (SN) in the LMC is SN1987A, as the only nearby core-collapse SN. SN1987A provides a perfect opportunity to study the

evolution of young Type II SNe into the SNR stage. It has therefore been the subject of several publications and observed by several instruments, including ATCA (Zanardo et al., 2013), XMM-Newton (Haberl et al., 2006), Chandra (Burrows et al., 2000) and recently JWST (Matsuura et al., 2024).

In this study we focus on Bayesian imaging methods for X-ray astronomy based on the algorithm D3PO (Selig and Enßlin, 2015), which implements denoising, deconvolution and decomposition of count data. Decomposition means that in addition to the total photon flux, the composition of the flux at each pixel is reconstructed using assumptions about prior statistics. The algorithm has been extended by Pumpe et al. (2018) to reconstruct and decompose multi-domain knowledge. The developed algorithm has been applied to Fermi data to reconstruct the spatio-spectral gamma-ray sky in Scheel-Platz et al. (2023) and its capabilities have been shown on X-ray photon-count data for Chandra in Westerkamp et al. (2024a). Here, we build a novel likelihood model for the eROSITA instrument and advance the prior model for the X-ray sky to reconstruct LMC features from EDR eROSITA data, as shown below. Moreover, we use variational inference (VI) to approximate the posterior instead of the maximum a posteriori (MAP) approach followed in Pumpe et al. (2018).

4.2 Observations

First draft written by Margret Westerkamp.

We have employed observations from the EDR of the LMC SN1987A containing data from the calibration and the performance verification phase of eROSITA. We have used data of the LMC SN1987A from eROSITA in pointing mode with the observation ID 700161.¹ In total, this observation of the LMC includes all seven telescope modules (TMs) of eROSITA. However, we have chosen only to use data from five of these TMs, specifically TM1, TM2, TM3, TM4 and TM6 (together usually referred to as TM8), since TM5 and TM7, as noted in Merloni et al. (2024), do not have an on-chip optical blocking filter and suffer from an optical light leak (Predehl et al., 2021). The raw data were processed using the eSASS pipeline (Brunner et al., 2022) and binned into 1024×1024 spatial bins and 3 energy bins, 0.2 - 1.0 keV (red), 1.0 - 2.0 keV (green) and 2.0 - 4.5 keV (blue), according to the binning used by Haberl et al.²

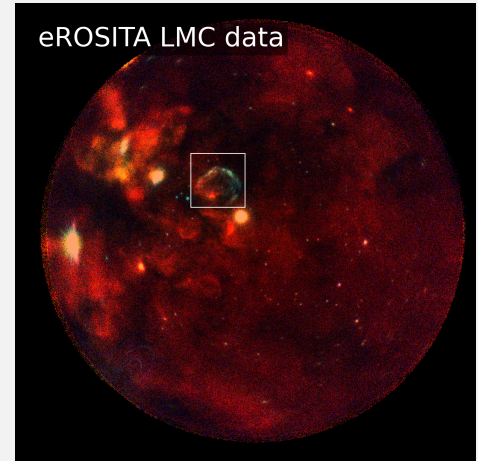


Figure 4.1: Visualisation of the exposure corrected eROSITA EDR TM8 data of ObsID 700161 for three energy bins. Red: 0.2 - 1.0 keV, Green: 1.0 - 2.0 keV, Blue: 2.0 - 4.5 keV. The white box marks the region of 30 Doradus C further discussed in section 4.3.2.

¹The data used are publicly available at <https://erosita.mpe.mpg.de/edr/eROSITA0bservations/>.

²The first light EDR image of LMC SN1987A by F. Haberl et al. is shown in https://www.dlr.de/de/aktuelles/nachrichten/2019/04/20191022_first-light-erosita.

In particular, we used `evtool` to generate the cleaned event files and `expmap` to generate the corresponding exposure maps for each TM. The specific configurations for `evtool` and `expmap` can be found in appendix C.1. We created new detector maps that included the bad pixels in order to exclude those from the inference. Figure 4.1 shows the corresponding RGB image of the eROSITA LMC data, where one image pixel corresponds to four arc seconds. In the appendix in fig. C.1 the data per energy bin and TM are shown. Figure C shows the exposures summed over the TMs.

4.3 Mathematical derivations

First draft written by Margret Westerkamp.

In this section, we present the methods used for X-ray imaging with eROSITA. In the end we want to reconstruct a signal s , in our case the X-ray photon flux density field in units of $[1/(\text{arcsec}^2 \times \text{s})]$. The signal is described by a physical field and is a function of spatial coordinates, $x \in \mathbb{R}^2$, and a spectral coordinates, $y = \log(\mathcal{E}/\mathcal{E}_0) \in \mathbb{R}$, where \mathcal{E} is the energy and \mathcal{E}_0 the reference energy. In the following, we describe the Bayesian inference of the signal field and its components, in other words the prior and the likelihood model.

4.3.1 Imaging with information field theory

First draft written by Matteo Guardiani.

X-ray imaging poses a series of different challenges. Astrophysical sources emit photons at a certain rate. This rate can be mathematically modeled by a scalar field which varies across the field of view (FOV), energy, and time. After being bent through the instrument's optics, this radiation is then collected by the charge-coupled devices (CCDs) which record individual photon counts as events. This way, the physical information contained in the sources' flux spatio-spectro-temporal distribution is degraded into the observational data. The mathematical object of a field with an infinite number of degrees of freedom, which is well suited to describe the original flux-rate signal, is not suited to describe a finite collection of event counts. Recovering the infinite degrees of freedom of the signal field from finite data is a challenging problem that requires additional information. IFT (Enßlin et al., 2009; Enßlin, 2019) provides the mathematical framework to introduce these additional components and solve the inverse problem of recovering fields from data. The additional information introduced characterizes typical source types found in astrophysical observations, such as point sources, which can be bright but are spatially sparse; diffuse emission, which is nearly ubiquitous across the FOV and spatially correlated; and extended sources, which are finite regions of diffuse emission with their own specific correlation structures. In the context of X-ray imaging, this allows to accurately and robustly reconstruct the underlying photon flux field as the sum of all modeled emission fields. In essence, upon denoting the quantity of interest, in our case the X-ray flux, with s for signal, we can use prior information on the distribution of s , $\mathcal{P}(s)$, to obtain posterior information $\mathcal{P}(s|d)$ on

the signal constrained to the observed data d using Bayes' theorem,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s) \mathcal{P}(s)}{\mathcal{P}(d)}. \quad (4.1)$$

Here, $\mathcal{P}(d|s)$ is called the likelihood and incorporates information about the instrument's response and noise statistics, while $\mathcal{P}(d)$ is called the evidence and ensures proper normalization of the posterior $\mathcal{P}(s|d)$. In the following, we will discuss our choices for the prior distribution (section 4.3.2), describe how to build the likelihood model which takes into account eROSITA-specific instrumental effects (section 4.3.3), and explain how to combine our likelihood and prior models to numerically approximate the posterior distribution as this turns out to be analytically intractable (section 4.3.4). The corresponding models are built using the software package J-UBIK (chapter 5), the JAX-accelerated universal Bayesian imaging kit, which is based on NIFTy.re (Edenhofer et al., 2024) as a JAX-accelerated version of NIFTy (Selig et al., 2013; Arras et al., 2019).

4.3.2 Prior models

First draft written by Margret Westerkamp.

Prior models are an essential part of Bayesian inference, allowing us to infer a field with a virtually infinite number of degrees of freedom from a finite number of data constraints. Here we explain how we mathematically model different sky components, the underlying assumptions and justifications, and how these models are implemented in a generative way. Our signal s is composed of a set of sky components $\{s_i\}$,

$$s = \sum_i s_i, \quad (4.2)$$

that differ in their morphology. In this study, these are in particular the point source emission, s_p , and the diffuse extended source emission, s_d . Building individual prior models for each of these components allows us to decompose the reconstructed, denoised and deconvolved sky into its various sources. The prior models for each of the sky components are implemented as generative models as introduced in Knollmüller and Enßlin (2020) using the reparametrization trick of Kingma et al. (2015). In other words, each of the prior models is described by a set of normal or log-normal models, leading to the final generative model defined via Gaussian processes via inverse transform sampling (see chapter 2 for 1D example). In this study, we distinguish between spatially correlated sources, which describe diffuse emission, and spatially uncorrelated sources, which model point sources. For each of the components we have a correlated spectral direction.

There are several ways to implement the correlation in the spatial or spectral dimension. To model the two-dimensional spatial correlation in diffuse emission, we use the correlated field model introduced in Arras et al. (2022). In this particular case the two-dimensional field, which we call $\varphi_{\text{ln}} = e^\tau$, is modelled by a log-normal process with τ being normal distributed, $\mathcal{P}(\tau|T) = \mathcal{N}(\tau, T)$, with unknown covariance T ,

$$\varphi_{\text{ln}} = e^\tau = e^{A\xi_\tau} \quad \text{with} \quad T = AA^\dagger, \quad \xi_\tau = \mathcal{N}(\xi_\tau, 1). \quad (4.3)$$

Here, a normal distribution is denoted as,

$$\mathcal{N}(x, X) := \frac{1}{|2\pi X|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x^\dagger X^{-1}x\right). \quad (4.4)$$

Assuming a-priori statistical homogeneity and isotropy, the correlation structure encoded in T can be represented by its power spectrum according to the Wiener-Khinchin theorem. In order to learn the power spectrum and thus the correlation structure simultaneously with the diffuse sky realization, it is implemented by an integrated Wiener process whose parameters are themselves represented by log-normal and Gaussian processes and can thus be learned from the data. For more details on the correlated field model see Arras et al. (2022).

For the point sources, we want the two-dimensional spatial field, φ_{ig} , to be pixel-wise uncorrelated, or in other words we want each pixel to be independent. Statistically this is described by probability distribution, which factorizes in spatial direction. Moreover, we aim for a few bright point sources. As shown in Guglielmetti et al. (2009) an appropriate probability distribution is the inverse gamma distribution, i.e.

$$\mathcal{P}(\varphi_{\text{ig}}) = \prod_x \Gamma^{-1}\left(\varphi_{\text{ig}}(x)\right). \quad (4.5)$$

As we aim to perform a spatio-spectral reconstruction of the eROSITA X-ray sky, we add a spectral axis. In this study we consider a power-law behavior, described by the spectral index α in spectral direction. For the diffuse emission we assume that the spectral index, α_d , is spatially correlated, while it is assumed to be spatially uncorrelated for the point source emission α_p . This leads to the mathematical definition of the individual components, s_p and s_d ;

$$s_p(x, y) = \varphi_{\text{ig}}(x) \times e^{\alpha_p(x)y} \quad \text{and} \quad s_d(x, y) = \varphi_{\text{ln}}(x) \times e^{\alpha_d(x)y}. \quad (4.6)$$

In fig. 4.1 it can be seen that both the correlation structure and the spectral power law behavior in the region of 30 Doradus C are fundamentally different from the diffuse structures that are otherwise present in the data. For the diffuse structures in the LMC we expect long correlation structures and a steep power-law slope in the energy direction. 30 Doradus C, on the other hand, has a flat power-law and a shorter correlation length. To account for this, we add another prior component, s_b , in the region of a box, b , around 30 Doradus C, which has a correspondingly flatter power law and allows for smaller structures, giving us a third component,

$$s_b(x, y) = \begin{cases} \varphi_{\text{ln}, b}(x) \times e^{\alpha_b(x)y} & \text{if } x \in b \\ 0 & \text{otherwise} \end{cases}. \quad (4.7)$$

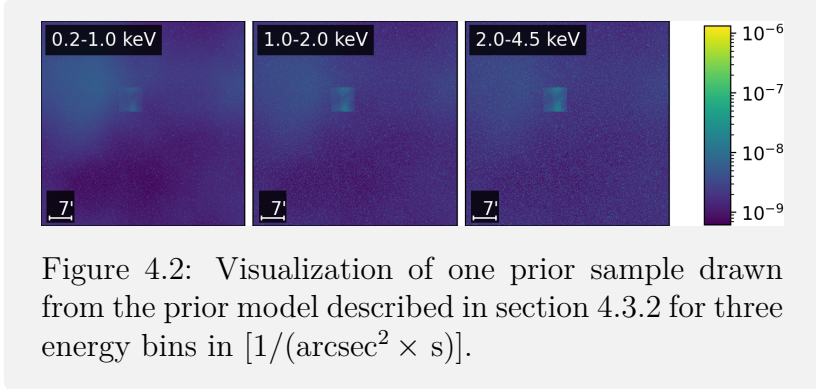


Figure 4.2: Visualization of one prior sample drawn from the prior model described in section 4.3.2 for three energy bins in $[1/(\text{arcsec}^2 \times \text{s})]$.

Figure 4.2 shows one prior sample drawn from the here described prior model for three energy bins. In the appendix C.2, we explain how to choose the latent parameters of this generative model in order to find a reasonable prior.

4.3.3 The likelihood

First draft written by Vincent Eberle.

The likelihood is the conditional probability of a data realization d given the underlying physical signal s . In the case of photon-count instruments like eROSITA, this conditional probability for a pixel i , takes the form of a Poisson distribution

$$\mathcal{P}(d_i|s) = \mathcal{P}(d_i|\lambda_i(s)) = \frac{\lambda_i^{d_i}}{d_i!} e^{-\lambda_i}, \quad (4.8)$$

with d_i being the photon counts and λ_i being the average number of photon counts on the detector pixel i , caused by the signal s . For a CCD chip with n instrument pixels the data is a vector of pixel photon counts, $d = (d_i)_{i \in \{1, \dots, n\}}$. The total likelihood turns into the product of the individual likelihoods in the case of statistical independence of the pixel events,

$$\mathcal{P}(d|s) = \prod_i \mathcal{P}(d_i|\lambda_i(s)) = \prod_i \frac{\lambda_i^{d_i}}{d_i!} e^{-\lambda_i}. \quad (4.9)$$

Often we refer to the negative logarithm of this probability as the likelihood information Hamiltonian

$$\mathcal{H}(d|s) = -\ln \mathcal{P}(d|\lambda(s)) = \sum_i \lambda_i - d_i \ln(\lambda_i) + \ln(d_i!). \quad (4.10)$$

These equations can be generalized to multiple observations m of the same sky with different instruments or at different times. Then the data is a vector of vectors, $d = (d_j)_{j \in \{1, \dots, m\}} = (d_{ji})_{j \in \{1, \dots, m\}, i \in \{1, \dots, n\}}$, where d_{ji} is the data point from the pixel i in the observation j which turns eq. (4.10) into

$$\mathcal{H}(d|s) = \sum_j \mathcal{H}(d_j|\lambda_j(s)) = \sum_{j,i} \lambda_{ji} - d_{ji} \ln(\lambda_{ji}) + \ln(d_{ji}!). \quad (4.11)$$

The steps performed to bin the data before using it in this formula are explained in section 4.2. In order to evaluate the Hamiltonian $\mathcal{H}(d|s)$ we need a digital representation of the measurement process, the relation between the physical signal s , and the mean count rate λ . The derivation of these quantities is discussed in this section.

Instrument model

An accurate description of the measurement process is essential for the inference of the signal s . Therefore, we need the instrument response \mathcal{R} , which represents the effects of the measurement process,

$$\lambda = \mathcal{R}(s), \quad (4.12)$$

to be as accurate as possible. However, since this function will be called many times during the computation process, it also has to be efficient and therefore we aim for a representation that is not only precise but also computationally affordable. In essence, we want to build a forward model that describes the linear effects of the measurement process. We tackle this by subdividing the response function \mathcal{R} into its most relevant constituents. The photon flux s coming from the sky gets smeared out by the PSF of the mirror assembly (MA). This gets mathematically described by an operator O . The PSF of each individual mirror module on-ground, on-axis and in-focus is of the order of 16.1 arcsec. However, the modules are mounted intra-focal to reduce the off-axis blurring for the price of an enlarged PSF in the on-axis region. Therefore, the in-flight PSF is significantly broader than 16.1 arcsec (Predehl et al., 2021). The blurred flux gets then collected by the camera assembly (CA). We denote the mathematical operator representing the exposure with E . It encodes the observation time and detector sensitivity effects. The flagging of invalid detector pixels, also called the mask, is denoted with M . The instrument response is thus

$$\mathcal{R} = M \circ E \circ O, \quad (4.13)$$

where \circ denotes the composition of operators. Readout streaks are almost completely suppressed due to the fast shift from the imaging to the frame-store area of the CCDs and therefore, don't have to be modeled (Predehl et al., 2021). Other effects, like pile-up are neglected up to this moment, but will be covered in future work. In the following sections, the parts of the instrument response will be discussed individually.

The point spread function

The PSF, here denoted as the mathematical operator O , describes the response of the instrument to a point-like source. An incoming photon from direction $\tilde{x} \in \mathbb{R}^2$ is deflected to a different direction $x \in \mathbb{R}^2$. This blurs the original incident flux s to the blurred flux s' , which is notated in a continuous and discretized way,

$$s'(x) = \int_{\mathbb{R}^2} O(x, \tilde{x}) s(\tilde{x}) d\tilde{x} \quad \text{or} \quad s'_x = \sum_{\tilde{x}} O_{x\tilde{x}} s_{\tilde{x}}. \quad (4.14)$$

This operator $O(x, \tilde{x})$ can be regarded as a probability density function $\mathcal{P}(x|\tilde{x})$, which is normalized by the integration over the space of all directions meaning, that the process of blurring conserves the photon flux,

$$1 = \int_{\mathbb{R}^2} O(x, \tilde{x}) dx = \int_{\mathbb{R}^2} \mathcal{P}(x|\tilde{x}) dx. \quad (4.15)$$

In the discretized form the operator O is a matrix and thus scales quadratically with the number of pixels n , resulting in a computational complexity of $\mathcal{O}(n^2)$. In most applications a spatially in-variant PSF is assumed, meaning that the PSF is the same for all points within the field of view. Thus the PSF is only a function of the deflection $x - \tilde{x}$, meaning $O(x, \tilde{x}) = O(x - \tilde{x})$. This fact turns eq. (4.14) into a convolution

$$s'(x) = \int_{\mathbb{R}^2} O(x - \tilde{x}) s(\tilde{x}) d\tilde{x}. \quad (4.16)$$

Convolutions on regular grids can be executed very efficiently, thanks to the convolution theorem and the fast Fourier transform (FFT) developed by Cooley and Tukey (1965). However, the assumption of spatial in-variance of the PSF only holds, depending on the variability of the PSF, for smaller FOVs. In order to image large structures on the sky, the spatial variability of the PSF cannot be ignored without imprinting artifacts on the reconstructions.

Therefore, we need a representation of spatially variant PSFs that can be used in the forward model. Here, we use the algorithm of (Nagy and O’Leary, 1997). This algorithm, which we call *linear patched convolution* in the following, is a method to approximate spatially variant PSFs in a computationally efficient way. It scales sub-quadratically, meaning it is computationally affordable, but improves the accuracy, in comparison to a regular spatially in-variant convolution.

In linear patched convolution the full spatially-variant PSF, O , is approximated by a combination of operations

$$O \approx \sum_k P_k W C_k. \quad (4.17)$$

First, the image is cut into k overlapping patches by the slicing operator C_k . Next, these patches are weighted with a linear interpolation kernel W , such that the total flux s , despite the overlapping patches, is conserved. Then, each patch is convolved with the associated PSF corresponding to the center of the patch, denoted by P_k . Finally, the results of the weighted and convolved overlapping patches are summed up. This can be seen as an Overlapp-Add convolution with linear interpolation and different PSFs for each patch (see (Nagy and O’Leary, 1997)).

In order to perform this operation we need information about the spatial distribution of the PSF, which we can retrieve from the calibration database (CALDB)¹. Here, we find information about the PSF, gathered at the PANTER 130 meter long-beam X-ray experimental facility of the Max-Planck-Institute for extraterrestrial physics (Predehl et al., 2021).² The CALDB files contain the measurements of the PSF for certain off-axis angles and energies, averaged over the azimuth angle. For the linear patched convolution algorithm at use we need the PSF at the central positions in the patches. To obtain these, we rotate and linearly interpolate the PSFs from the CALDB, which allows us to construct the PSFs at these central positions. We also remove some noticeable shot noise from the measured PSFs by clipping the normalized PSFs at 10^{-6} .

¹Information about the CALDB: <https://erosita.mpe.mpg.de/edr/DataAnalysis/esasscaldb.html>

²Details in appendix C.1

The exposure

The received flux on the camera is observed for a total exposure time Θ by the CA. The exposure operator E includes not only the exposure time Θ , but also the vignetting, ρ , of the TM, and its effective area A_{eff} . In the case of a time-invariant flux $s'(t) = s'_0$, the integral over time corresponds to a multiplication with the total exposure time Θ and thus

$$\lambda = \int_{\Theta} \rho A_{\text{eff}} s'(t) dt = \rho A_{\text{eff}} s'_0 \int_{\Theta} dt = (\rho A_{\text{eff}} \Theta) s'_0 = E s'_0. \quad (4.18)$$

In the case that $s'(t)$ is not constant, s'_0 is the average value of $s'(t)$ in the observed time interval. We calculate the total observation time of a pixel projected to the sky in a certain energy band, combined with the vignetting, with **eSASS** (eROSITA Science Analysis Software System)³ (Brunner et al., 2018; Predehl et al., 2021) through the command **expmap**. The parameters used for the **expmap** command can be found in appendix C.1 in table C.2. The information about the effective area for each TM can be found in the **CALDB**⁴.

The mask

The likelihood information Hamiltonian eq. (4.11) derived from the Poissonian distribution is only defined for $\lambda > 0$. Also it contains the logarithm of the count rate λ . Therefore, it is necessary to mask all sky positions with no observation time or defective detector pixels for a certain TM, which result in $\lambda = 0$ for the respective pixel.

Removing these pixels from the calculation makes the algorithm more stable, prevents the appearance of **NaNs**, and ensures that only reliable data is used for the reconstruction. From the raw data, there seem to be corrupted data points in regions with very low exposure time and at the boundary of the FOV. Therefore, we decided to mask from the reconstruction all pixels and data points with an observation time of less than 500 seconds.

As not all bad pixels were correctly flagged in the exposure files, we used information from the **CALDB** badpix files to update the detmap files. We then used these modified detmap files to build the new exposure maps and update the mask.

Forward model for multiple observations

In the case of an eROSITA pointing, all TMs that are online observe the same sky and capture the same physics. Although the instruments are very similar, they are not identical. Their slightly different pointing results in different positions of the focal point of the PSF. Also, they may have different good-time intervals, resulting in different exposure times and also different defective pixels for the CCDs. Instead of summing the counts from the different data sets, thereby assuming a “mean” instrument, we model each TM and its observation individually. That means, we formulate the signal response λ_j of one TM_j as

$$\lambda_j = M_j E_j O_j s. \quad (4.19)$$

³More information about the **eSASS** software developed by the eROSITA Team can be found here: <https://erosita.mpe.mpg.de/edr/DataAnalysis/>

⁴Details in appendix C.1

We display a visualization of the forward model’s computational graph in fig. 4.3. By plugging in all λ_j into eq. (4.11), we get a formulation of the full likelihood information Hamiltonian that allows us to remove the individual detector effects jointly.

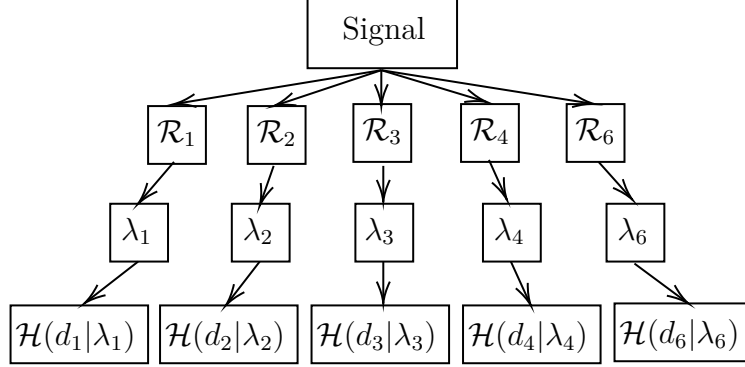


Figure 4.3: Visualisation of the computational graph of the forward model.

4.3.4 Inference

First draft written by Matteo Guardiani.

In principle, given the prior and likelihood distributions, eq. (4.1) allows to fully determine the posterior distribution by computing the evidence

$$\mathcal{P}(d) = \int_{\Omega_s} \mathcal{D}s \, \mathcal{P}(d|s) \mathcal{P}(s),$$

where we have denoted with Ω_s the Hilbert space in which s lives. In general, and specifically for the prior and likelihood models described above, the evidence cannot be explicitly evaluated, as it would require integrating over the potentially multi-million- or multi-billion-dimensional space, Ω_s . To overcome this problem we use VI. In VI, the evidence calculation problem is overcome by approximating the posterior distribution directly using a family of tractable distributions $\mathcal{Q}_\phi(s|d)$, parametrized by some variational parameters ϕ . To approximate the posterior we minimize the Kullback–Leibler divergence (KL),

$$\mathcal{D}_{\text{KL}}(\mathcal{Q}_\phi||\mathcal{P}) := \int_{\Omega_s} \mathcal{D}s \, \mathcal{Q}_\phi(s|d) \log \frac{\mathcal{Q}_\phi(s|d)}{\mathcal{P}(s|d)},$$

with respect to the variational parameters ϕ . In this work, the family of approximating posterior distributions $\mathcal{Q}_\phi(s|d)$ is built using geoVI (Frank et al., 2021). In geoVI, the posterior is approximated with a Gaussian distribution in a space in which the posterior is approximately Gaussian. This is achieved by utilizing the Fisher information metric, which captures the curvature of the likelihood and the prior distributions. The Fisher metric provides a way to measure the local geometry of the posterior, guiding the creation of a local isometry – a transformation that maps the curved parameter space to

a Euclidean space while preserving its geometric properties. In this transformed space, the posterior distribution approximates a Gaussian distribution more closely, allowing the Gaussian variational approximation to be more accurate. Consequently, geoVI can represent non-Gaussian posteriors with high fidelity, improving inference results. By leveraging the geometric properties of the posterior distribution, geoVI offers a powerful extension to traditional VI, enabling more precise and reliable approximations for complex Bayesian models, as the ones presented in this work.

4.4 Results

First draft written by Matteo Guardiani.

In fig. 4.4, we present the reconstruction of the sky flux distribution based on the data shown in section 4.2. Our algorithm's forward modeling of the X-ray sky enables the decomposition of the signal into point-like, diffuse, and extended-source emission components, providing a more detailed view of the small-scale features of the extended structure of the 30 Doradus C bubble. These reconstructed components are also displayed in fig. 4.4. From these reconstructions, it is clear that most of the point-like emission is well separated from the diffuse emission, resulting in the first denoised and deconvolved view of this region of the sky as observed by the eROSITA X-ray observatory. Additionally, in fig. C.5, we show the reconstructed flux for each energy bin, offering a clearer understanding of the color scheme adopted in fig. 4.4. All the final reconstructions have been obtained using the geoVI algorithm. For the spatial distribution, we have chosen a resolution of 1024×1024 pixels. For the spectral distribution, we have chosen 3 energy bins corresponding to the energy ranges between 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV, respectively. The variational approximation to the posterior was estimated using 8 samples, corresponding to 4 pairs of antithetic samples. We considered the posterior approximation to have converged when the posterior expectation values of the signals of interest, such as the reconstructed sky flux field, exhibited no significant changes between consecutive iterations of the VI algorithm. The runtime for the reconstruction was approximately one day on a CPU for a single module, and around two days for all five analyzed telescope modules. By adopting a fully probabilistic approach, we leverage posterior samples to assess how well the model assumptions align with the observed data. In particular, in the presence of shot noise, we define the NWR as

$$\epsilon_{\text{NWR}} = \left\langle \frac{\lambda(s) - d}{\sqrt{\lambda(s)}} \right\rangle_{\phi} \simeq \frac{1}{N} \sum_{i=0}^N \frac{\lambda(s_i^*) - d}{\sqrt{\lambda(s_i^*)}}, \quad (4.20)$$

where $\lambda(s)$ is the expected number of counts predicted by the model and d is the observed data. Here, the posterior average $\langle \cdot \rangle_{\phi}$ over \mathcal{Q}_{ϕ} is approximated by the sample average over the corresponding posterior samples s_i^* . These residuals are particularly useful for identifying model inconsistencies, which may indicate areas for improving the instrument's description as well as point to potential calibration improvements. We will explore this possibility further in section 4.5.

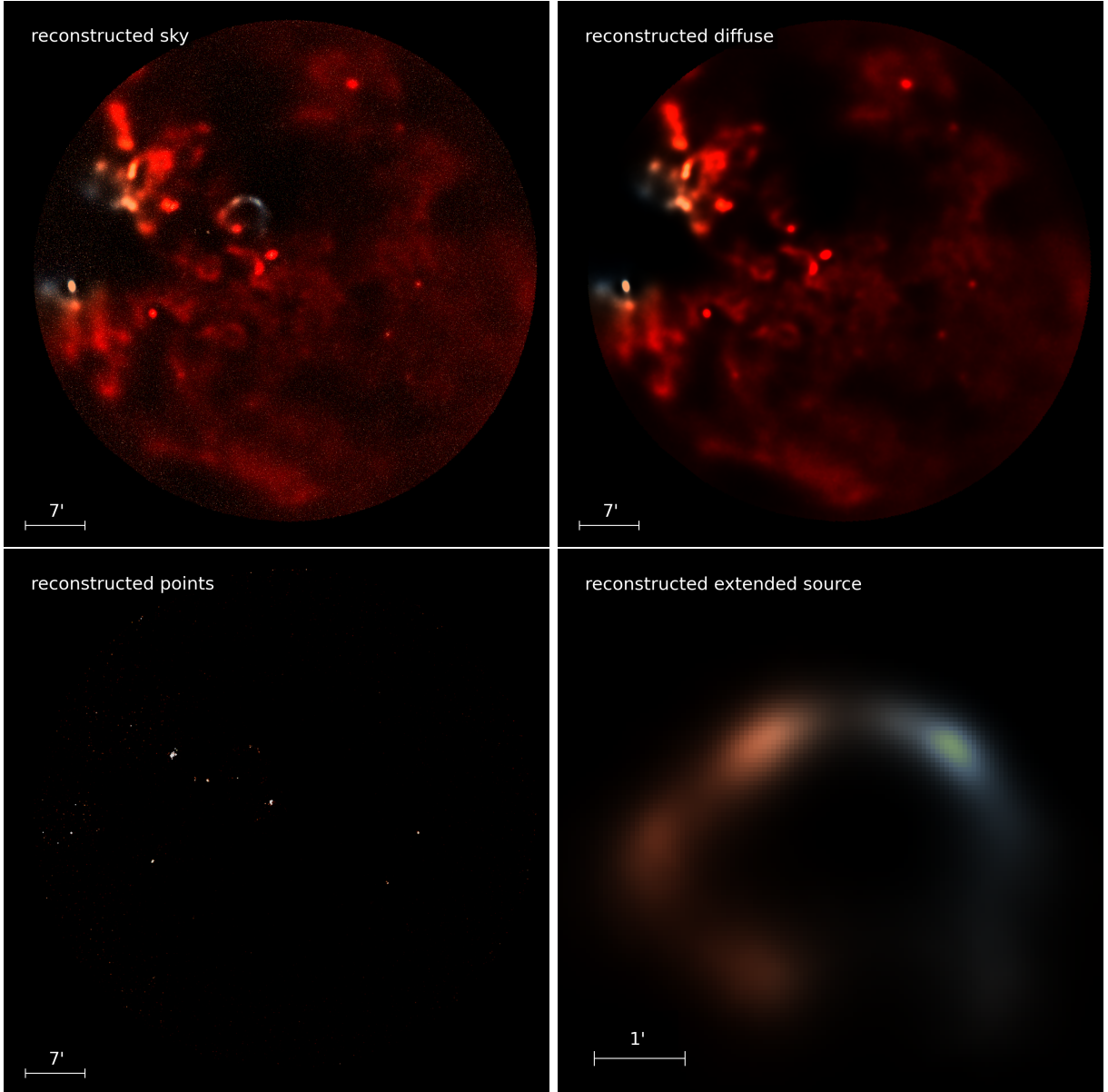


Figure 4.4: Posterior mean of the SN1987A reconstruction. *The top panels* display on the left the reconstruction of the sky and on the right the separated diffuse emission. *The bottom panels* display the reconstruction of the point-like emission (left) and the reconstruction of the diffuse emission from 30 Doradus C (right). We convolve the point sources with an unnormalized Gaussian kernel with standard deviation $\sigma = 0.5$, in order to make them visible on printed paper. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively.

4.5 Discussion and validation

First draft written by Margret Westerkamp.

The discussion is divided into two parts. First, in section 4.5.1, we validate the general consistency of the presented algorithm using a simulated sky and simulated data, which also motivates the detection threshold for point sources. In the second part, section 4.5.2, we discuss the results of the reconstruction presented in section 4.4 along with corresponding diagnostics, such as the NWRs.

4.5.1 Validation

Generative modeling allows to generate prior models of the sky, as described in section 4.3.2. These prior models can be used to validate the consistency of the presented algorithm. In particular, we look at prior samples of the X-ray sky, composed out of point sources and diffuse emission, with a FOV of 1024 arcsec. Using the same resolution as for the actual reconstruction this leads to 256×256 pixels. We pass the prior samples through the forward model shown in fig. 4.3 including all five TMs, which gives us simulated data. Figure 4.5 shows the considered prior sample of the X-ray sky as well as the corresponding simulated data passed through the eROSITA response and affected by Poissonian noise. The simulated data per TM and the underlying simulated sky per energy bin is shown in the appendix C.4 in fig. C.9 and fig. C.10. Using the simulated data, we aim to apply the algorithm presented above to estimate the posterior through VI posterior samples. We then evaluate how well the corresponding simulated sky, or prior sample, is reconstructed and determine the corresponding uncertainty in that estimate. The right side of fig. 4.5 shows the reconstructed prior sample. Component separation, deconvolution, and denoising techniques show strong performance when applied to simulated data. They effectively recover the underlying signal.

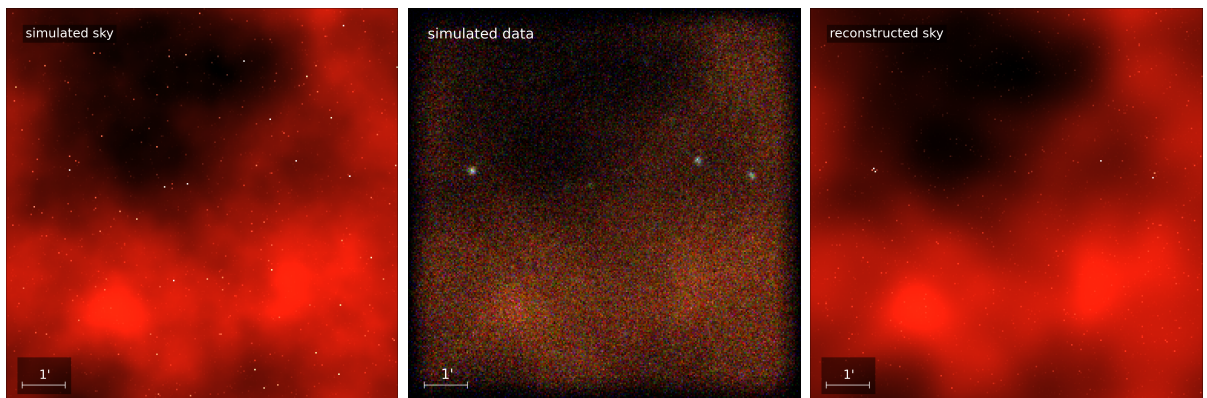


Figure 4.5: Visualisation of the simulated X-ray, the corresponding simulated data and its reconstruction. Left: Simulated X-ray sky. Center: Simulated X-ray data generated as shown in fig. 4.3. Right: Reconstructed X-ray sky.

To validate the results shown in fig. 4.5, we use a set of validation metrics that we have access to due to the probabilistic approach of the algorithm. These metrics are intended to provide further insight into the residuals between the simulated X-ray sky and its reconstruction, as well as the uncertainty of the algorithm at each pixel.

Accordingly, we show in the appendix the standard deviation of the posterior samples in fig. C.12, which gives us a measure of the uncertainty of the algorithm. To examine the residuals, we define the standardized error as the relative residual between the ground truth, s_{gt} , and the posterior mean, μ_s ,

$$\epsilon_{\text{rel}}(s_{\text{gt}}) = \frac{\mu_s - s_{\text{gt}}}{s_{\text{gt}}}, \quad (4.21)$$

to check for differences between the ground truth and the reconstruction. This standardized error is shown for each energy bin in appendix C.4 in fig. C.13. The image shows that point sources are not detected or are misplaced in some areas. This highlights the need for a detection threshold for point sources in the reconstruction to ensure the correctness as also indicated in the hyper parameter search in appendix C.2. To validate the detection threshold further we use posterior samples, s_p^* , of the approximated posterior, $\mathcal{Q}_\phi(s_p|d)$, for the point source component in order to get the sample-averaged two-dimensional histogram of the absolute standardized error only for point sources, $|\epsilon_{\text{rel}}^*|(s_{\text{gt}, p})$, where,

$$|\epsilon_{\text{rel}}^*|(s_{\text{gt}, p}) = \frac{|s_p^* - s_{\text{gt}, p}|}{s_{\text{gt}, p}}. \quad (4.22)$$

Figure 4.6 shows the sample-averaged histogram together with the detection threshold, θ , analytically set for this reconstruction in appendix C.2. Below the detection threshold, the histogram shows two effects, undetected or misplaced point sources and possible noise over-fitting, which are eliminated by cutting the point sources below the detection threshold to ensure the consistency of the reconstruction. We apply the same cuts to the reconstruction shown in section 4.4.

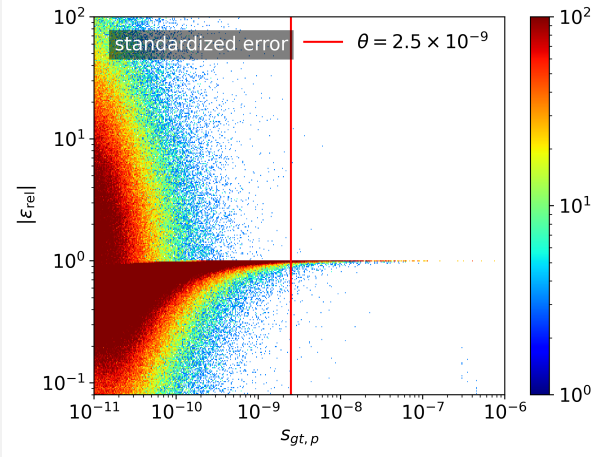


Figure 4.6: Two-dimensional histogram of the standardized error eq. (4.22) for the point sources. The histogram is plotted together with the detection threshold, $\theta = 2.5 \times 10^{-9}$, calculated in appendix C.2.

4.5.2 Discussion of results

The results of the algorithm described above, applied to the eROSITA LMC data, are shown in section 4.4. Figure 4.4 shows the LMC in a deconvolved, denoised and decomposed view. The full image of the LMC is shown, as well as the separated components of the point sources, the diffuse structures of the LMC, and the extended sources of 30 Doradus C. As a result of the inference we get posterior samples of the approximated posterior probability $\mathcal{Q}_\phi(s|d)$. Given these posterior samples, we can calculate a measure of uncertainty of the reconstruction, which is in this case given by the standard deviation. The corresponding plots of the standard deviation per energy bin for the reconstruction shown in fig. 4.4 are shown in appendix C.3 in fig. C.6. As expected, we can see that the uncertainty is higher in regions of a high number of photon counts.

Analyzing the component separation in fig. 4.4, it can be seen that there is still a halo around the source of SN1987A, which can have two different causes, firstly a detection pile-up effect due to the high fluxes from these sources (Davis, 2001), and secondly a mismodeling of the instruments due to calibration mismatches. In order to check for possible calibration issues we performed single-TM reconstructions, which only took the data and the response functions for one of the TMs each into account. The results of the single-TM reconstructions are shown per energy bin in appendix C.3 in fig. C.3. These images give us a great insight into possible calibration inconsistencies together with the corresponding NWRs (eq. (4.20)) per TM and energy bin, which are shown in fig. C.8. In particular, the reconstruction for TM2 suggests both pile-up issues and mismatches in the calibration files, such as the PSF and dead pixels. Although we incorporated information about the dead pixels into the inference, the number of dead pixels accounted for does not seem to be sufficient. The reconstruction clearly indicates that there are likely additional dead pixels in this area.

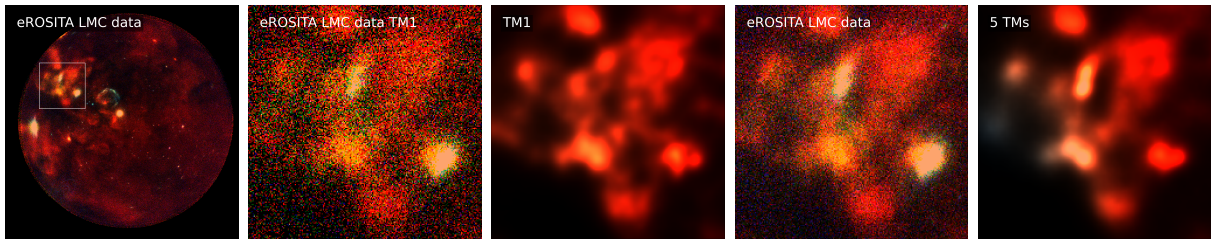


Figure 4.7: Zoom-in on reconstruction of diffuse emission from the Tarantula Nebula. From left to right showing the zoom area on the plot of the eROSITA LMC data, the zoomed LMC data for TM1, the corresponding single-TM reconstruction for TM1, the zoomed data for all 5 TMs and the reconstruction by means of all five observations.

If we look at the single-TM reconstruction and compare it with the reconstruction using all five TMs in fig. 4.4, we see that in the latter the diffuse structures of the LMC and 30 Doradus C have been much more finely resolved. Figure 4.7 shows a comparison of the diffuse structures around the Tarantula Nebula in the eROSITA data, zoomed in for both a single TM and all five TMs. This higher resolution is due to the higher amount of

information we get on the diffuse structure by incorporating several observations. However, we also find that it is more difficult to separate point-like from diffuse emission using all the telescope modules. This is likely due to possible calibration inconsistencies in comparison with the single-TM reconstructions.

4.6 Conclusion

First draft written by Margret Westerkamp.

In conclusion, this paper presents the first Bayesian reconstruction of the eROSITA EDR data, providing a denoised, deconvolved, and separated view of the diffuse and point-like sources in the LMC. The presented algorithm enables the spatio-spectral reconstruction of the LMC, incorporating its observation by the five different TMs of TM8. Ultimately, this reconstruction has the potential to assist in the further analysis of the diffuse X-ray emission as done by Sasaki et al. (2022) without any noise or point source contributions or effects from the PSF. It also allows the point source catalog to be refined by considering only the point source component. Due to the generative nature of the algorithm we are able to generate simulated data, on which we tested the consistency of the reconstruction. The underlying building blocks of the implementation are publicly available (Eberle et al., 2024b) and can therefore be used to image other eROSITA observations as well.

In particular, the presented algorithm uses an additional component in the region of 30 Doradus C. Such additional components in certain regions allow to image such extended objects which overlap with the emission from the hot phase of the ISM and point sources and have a very different correlation structure. In this way, not only the general diffuse and point source emission can be decomposed, but also the diffuse emission from the hot phase of the ISM and from extended sources such as 30 Doradus C. In this work, the additional component for the extended source was set by hand. For future work, we aim to automate this and to find the the extended sources for high excitations in the latent space.

There are also several areas for further investigation. The algorithm presented here can be useful to check the calibration using single TM reconstructions and diagnostics such as the NWRs, which are readily available due to the algorithm's probabilistic nature. Future work could focus on improving the spectral resolution to allow further insight into the spectra of the different components. In addition, work is underway to extend the applicability to eROSITA field scans and all-sky surveys. Ultimately, the goal is to use this algorithm to perform multi-instrument Bayesian reconstructions of specific celestial objects.

Chapter 5

J-UBIK: The JAX-accelerated Universal Bayesian Imaging Kit

*The following chapter is a manuscript that has been submitted to **The Journal of Open Source Software** with Vincent Eberle, Matteo Guardiani and me as the shared first authors (Eberle et al., 2024b).¹ The open source software package is the result of a close collaboration between Vincent Eberle, Matteo Guardiani and myself. The project benefited greatly from the knowledge and discussions with Philipp Frank regarding the inference structure and JAX implementation. The implementation of the JWST instrument and demo is the contribution of Julian Rüstig. Julia Stadler contributed the data loading pipeline for Chandra. The project was supported by many fruitful discussions with Torsten Ensslin and is based on his vision of a universal Bayesian imaging kit. The initial draft of the manuscript was written by me, forming the basis for the collaborative revisions that followed.*

For consistency within this thesis, some of the parameters have been renamed and some of the figures have been adapted according to the layout.

¹Further information on code and contributions can be found at <https://github.com/NIFTy-PPL/J-UBIK>

Many advances in astronomy and astrophysics originate from accurate images of the sky emission across multiple wavelengths. This often requires reconstructing spatially and spectrally correlated signals detected from multiple instruments. To facilitate the high-fidelity imaging of these signals, we introduce the universal Bayesian imaging kit (UBIK). Specifically, we present J-UBIK, a flexible and modular implementation leveraging the JAX-accelerated `NIFTY.re` software as its backend. J-UBIK streamlines the implementation of the key Bayesian inference components, providing for all the necessary steps of Bayesian imaging pipelines. First, it provides adaptable prior models for different sky realizations. Second, it includes likelihood models tailored to specific instruments. So far, the package includes three instruments: Chandra and eROSITA for X-ray observations, and the James Webb Space Telescope (JWST) for the near- and mid-infrared. The aim is to expand this set in the future. Third, these models can be integrated with various inference and optimization schemes, such as maximum a posteriori estimation and variational inference. Explicit demos show how to integrate the individual modules into a full analysis pipeline. Overall, J-UBIK enables efficient generation of high-fidelity images via Bayesian pipelines that can be tailored to specific research objectives.

5.1 Summary

Many advances in astronomy and astrophysics originate from accurate images of the sky emission across multiple wavelengths. This often requires reconstructing spatially and spectrally correlated signals detected from multiple instruments. To facilitate the high-fidelity imaging of these signals, we introduce the universal Bayesian imaging kit (UBIK). Specifically, we present J-UBIK, a flexible and modular implementation leveraging the JAX-accelerated `NIFTY.re` (Edenhofer et al., 2024) software as its backend. J-UBIK streamlines the implementation of the key Bayesian inference components, providing for all the necessary steps of Bayesian imaging pipelines. First, it provides adaptable prior models for different sky realizations. Second, it includes likelihood models tailored to specific instruments. So far, the package includes three instruments: Chandra and eROSITA for X-ray observations, and the James Webb Space Telescope (JWST) for the near- and mid-infrared. The aim is to expand this set in the future. Third, these models can be integrated with various inference and optimization schemes, such as maximum a posteriori estimation and variational inference. Explicit demos show how to integrate the individual modules into a full analysis pipeline. Overall, J-UBIK enables efficient generation of high-fidelity images via Bayesian pipelines that can be tailored to specific research objectives.

5.2 Statement of need

In astrophysical imaging, we often encounter high-dimensional signals that vary across space, time, and energy. The new generation of telescopes in astronomy offers exciting opportunities to capture these signals but also presents significant challenges in extracting the most information from the resulting data. These challenges include accurately modeling the instrument’s response to the signal, accounting for complex noise structures, and separating overlapping signals of distinct physical origin.

Here, we introduce J-UBIK, the JAX-accelerated Universal Bayesian Imaging Kit, which leverages Bayesian statistics to reconstruct complex signals. In particular, we envision its application in the context of multi-instrument data in astronomy and also other fields such as medical imaging. J-UBIK is built on information field theory IFT (Enßlin, 2013) and the NIFTy.re software package (Edenhofer et al., 2024), a JAX-accelerated version of NIFTy (Selig et al., 2013; Steininger et al., 2019; Arras et al., 2019).

Following the NIFTy paradigm, J-UBIK employs a generative prior model that encodes assumptions about the signal before incorporating any data, and a likelihood model that describes the measurements, including the responses of multiple instruments and noise statistics. Built on NIFTy.re, J-UBIK supports adaptive and distributed representations of high-dimensional physical signal fields and accelerates their inference from observational data using advanced Bayesian algorithms. These include maximum a posteriori (MAP), Hamiltonian Monte Carlo (HMC), and two variational inference techniques: metric Gaussian variational inference MGVI (Knollmüller and Enßlin, 2020; Enßlin, 2013) and geometric variational inference geoVI, (Frank et al., 2021). As NIFTy.re is fully implemented in JAX, J-UBIK benefits from accelerated inference through parallel computing on clusters or GPUs.

Building generative models with NIFTy.re for specific instruments and applications can be very tedious and labor-intensive. Here, J-UBIK comes into play which addresses this challenge from two angles. First, it provides tools to simplify the creation of new likelihood and prior models and acts as a flexible toolbox. It implements a variety of generic response functions, such as spatially-varying point-spread functions (PSFs) (Eberle et al., 2023) and enables the user to define diverse correlation structures for various sky components. Second, J-UBIK includes implementations for several instruments.

Currently, it supports Chandra, eROSITA pointings, and JWST observations, with plans to expand this list as the user base grows. This expansion will provide users with a diverse set of accessible inference algorithms for various instruments. Ultimately J-UBIK enables the user, through Bayesian statistics, not only to obtain posterior samples and hence measures of interest such as the posterior mean and uncertainty of the signal for a several data sets, but also to perform multi-instrument reconstructions.

The software has already been applied in chapter 3, and publications on eROSITA pointings and JWST are currently in preparation. In the future, the set of instruments will be further expanded to include existing imaging pipelines from NIFTy and NIFTy.re such as those described in Scheel-Platz et al. (2023), Roth et al. (2023), Hutschenreuter et al. (2022), as well as new ones.

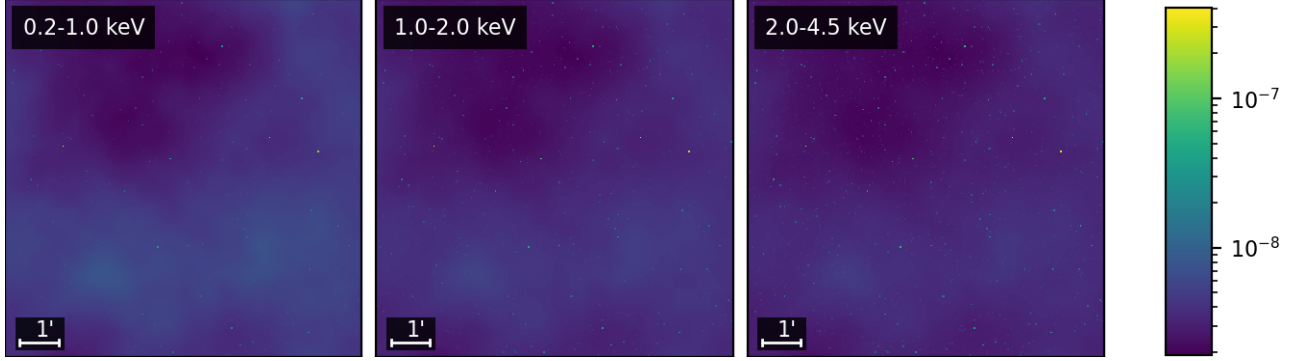


Figure 5.1: Simulated X-ray sky

5.3 Bayesian imaging with J-UBIK

At the core of the J-UBIK package is Bayes' theorem:

$$\mathcal{P}(s|d) \propto \mathcal{P}(d|s)\mathcal{P}(s), \quad (5.1)$$

where the prior $\mathcal{P}(s)$ represents our knowledge about the signal s before observing the data d , and the likelihood $\mathcal{P}(d|s)$ describes the measurement process. The posterior $\mathcal{P}(s|d)$ is the primary measure of interest in the inference process. J-UBIK's main role is to model the prior in a generative fashion and to facilitate the creation and use of instrument models to develop the likelihood model. The package includes demos for Chandra, eROSITA pointings, and JWST, which illustrate how to use or build these models and how to construct an inference pipeline to obtain posterior estimates.

5.3.1 Prior models

The package includes a prior model for the sky's brightness distribution across different wavelengths, which can be customized to meet user needs in both spatial and spectral dimensions. This model allows for the generation of spatially uncorrelated point sources or spatially correlated extended sources, as described by the correlated field model in Arras et al. (2022). In the spectral dimension, the model can be a power law, describe the correlation structure of the logarithmic flux using a Wiener process along the spectral axis or combine both of these models. The prior model's structure is designed to be flexible, allowing for modifications to accommodate additional dimensions and correlation structures. Figure 5.1 illustrates an example of a simulated X-ray sky in J-UBIK, sampled from a corresponding generative prior model with one energy bin. This example features two components: one representing spatially uncorrelated point sources and the other representing spatially correlated extended structures. Figure 5.1 shows from left to right the full sky and its components, the diffuse, extended structures and the point sources.

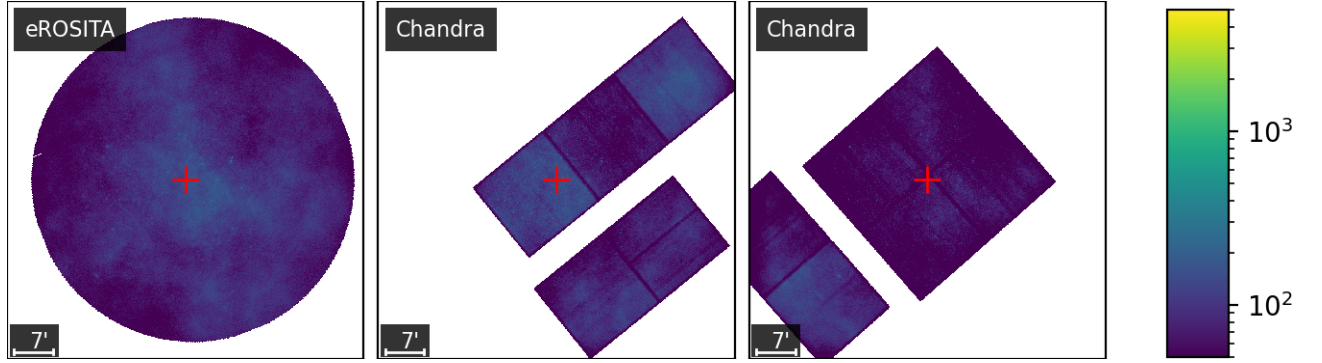


Figure 5.2: Simulated X-ray data

5.3.2 Likelihood models

J-UBIK implements several instrument models (Chandra, eROSITA, JWST) and their respective data- and response-loading functionalities, enabling their seamless integration into the inference pipeline. Due to its fully modular structure, we anticipate the inclusion of more instruments into the J-UBIK platform in the future. J-UBIK is not only capable of reconstructing signals from real data; since each instrument model acts as a digital twin of the corresponding instrument, it can also be used to generate simulated data by passing sky prior models through the instrument’s response. This allows to test the consistency of the implemented models.

Figure 5.2 shows the same simulated sky from fig. 5.1 seen by two different instruments, eROSITA and Chandra, with Poisson noise on the photon count data. The pointing center for each observation is marked in red. The two images on the right illustrate the same simulated sky seen by Chandra, but with different pointing centers, showing the impact of spatially varying PSFs (Eberle et al., 2023).

Chapter 6

Conclusion

Throughout this thesis, the common goal is to leverage the observed data to its fullest potential, whether in model comparison or in imaging. This work covers the formulation, implementation and application of field inference algorithms in a generative manner to exploit correlation structures and morphologies of the underlying, possibly superimposed, signal.

The main focus of this thesis is the reconstruction of the X-ray sky on the basis of photon event data. Three device-specific challenges are addressed by the likelihood model: the deconvolution, the denoising and the mosaicing of several overlapping data sets. The separation of multiple X-ray emitting sources and the inference of complex correlation structures of such components is driven by implemented generative and non-parametric prior models. As part of this work, the first Bayesian spatio-spectral prior model for the X-ray sky is implemented, which accounts for non-trivial correlation structures in the spatial and spectral directions. To speed up the reconstruction with the parameter-heavy spatio-spectral prior, both a novel transition model is proposed and the codebase is developed using JAX. Two implementations of likelihood and prior models for X-ray telescopes are demonstrated in this work: Chandra and eROSITA. The Bayesian field inference pipeline for Chandra, using a spatially invariant point spread function, is applied to eleven datasets of the supernova remnant of SN1006. The results reveal a distinct view of the resolved diffuse structures of the remnant and the sharp point sources. Intensity profiles along the shock front show strong and sharp X-ray flux peaks of up to two orders of magnitude. The eROSITA likelihood employs a spatially variant point spread function. The prior model describes three signal components: diffuse emission, point sources, and extended sources. Correspondingly, the inference results on the eROSITA EDR data of the LMC SN1987A show the diffuse emission, the point sources and 30 Doradus C separated. The reconstructed components show reduced noise and point spread function effects.

The software package **NIFTy** provides the numerical framework for IFT. Still, the construction of likelihood and prior models in **NIFTy** is a labor-intensive task. To support their implementation, the software package **J-UBIK** is developed as part of this thesis. **J-UBIK** aims to facilitate the implementation of future likelihood and prior models and to incorporate a growing set of instrument implementations - in its current state Chandra, eROSITA

and JWST. The result is a flexible and modular framework that supports Bayesian imaging as well as related diagnostic measures and validity checks.

Variational inference is used to approximate the posterior in the X-ray imaging algorithm, as described above. The thesis highlights, that in this particular case, the posterior normalization, the evidence, is not required. But when it comes to comparing different Bayesian models, it is important to calculate the evidence. Thus, the methodological part of this thesis evolves the formulation of a field inference problem for the evidence calculation using nested sampling. In nested sampling the lack of knowledge about the precise values of the prior mass introduces a stochastic error, which has a major impact on the evidence calculation. The present work leverages Bayesian field inference to infer the likelihood-prior-volume function. In particular, the assumptions of monotonicity and smoothness of the likelihood-prior-volume function are included in the inference. While standard nested sampling requires a trade-off between computational time and estimation accuracy, the proposed approach, serves as a post-processing step with a computational cost that remains independent of the number of live points. Thus, without a significant increase in computing time, the accuracy of the evidence calculation can be improved. The results show a significant increase in accuracy in the evidence calculation for less than hundred prior samples in the iterative process of nested sampling.

These findings lay the groundwork for future research in this field. The Bayesian field inference methods presented for evidence calculation and X-ray imaging mark initial steps, and accordingly, numerous challenges and further research directions remain to be addressed.

In this way, the presented post-processing for nested sampling holds the potential to improve the evidence calculation without the drawback of an exploding computational cost. Still, the algorithm encounters instability issues due to the Gaussian approximation of the δ -function in the theoretical framework. Thus, alternative approaches, such as those discussed in the appendix, could be explored further. Also, the method has its limits when dealing with non-smooth likelihood-prior-volume functions or plateaus. One approach discussed is to split the data set into several parts and apply post-processing only to the strictly monotonic parts. As this case demonstrates, future work on more evolved likelihood-prior-volume relations is desirable.

In the context of X-ray imaging, future research could focus on the auto-detection of extended sources, the refinement of the spectral resolution or implementation and incorporation of further instruments. Until now, as in the eROSITA imaging of the LMC SN1987A, extended sources have been introduced manually; automating their detection and extraction would be a significant advancement for future imaging. An enhanced spectral resolution could enable the separation of different emission processes. For example, by defining different prior models for thermal and non-thermal emission for the supernova remnant SN1006 one could imagine to decompose its spectral components. This paves the way for an evolved analysis of the spectra and the conditions within the region under consideration. As part of the J-UBIK package future implementations could address a growing database of instruments. These could include XMM-Newton or different observation modes of currently deployed instruments, like the eROSITA scans and all-sky

surveys. Cross-calibration and multi-instrument reconstructions within J-UBIK would be the ultimate goal.

In summary, this work contributes to the development and exploration of accelerated Bayesian inference techniques. It provides robust, open-source tools and methods that improve both the accuracy and interpretability of results in nested sampling and X-ray imaging.

Appendix A

Supplementary material chapter 2

A.1 Inference with Gaussian approximation

In the following, we want to work out an exact inference algorithm that is not based on a Gaussian approximation of the δ -function. This approach has its difficulties due to the calculation of the Fisher metric, as we will see below, and therefore cannot be used for geoVI. The approach discussed in the main part of this paper focuses on the joint reconstruction of the likelihood–prior–volume function and the prior volumes. The approach presented here solely reconstructs the likelihood–prior–volume function, given the data on the likelihood contours. The corresponding prior volumes for each likelihood contour can then be computed by inversion of this function. Again, we incorporate the smoothness assumption into the inference, but this time using a different relationship between the likelihood–prior–volume function and the correlated field,

$$\frac{df(\ln L)}{d \ln X} = -e^{\tau(f(\ln L))}. \quad (\text{A.1})$$

The solution for the prior volumes given the likelihood information is,

$$X(f(\ln L)|\tau) = \exp\left(\ln X_0 - \int_{f(\ln L_0)}^{f(\ln L)} e^{-\tau(y)} dy\right) = \exp\left(- \int_{f(\ln L_0)}^{f(\ln L)} e^{-\tau(y)} dy\right), \quad (\text{A.2})$$

with $X_0 = 1$. The prior model is then given by $\mathcal{P}(\tau)$, and our goal is to determine the likelihood model, denoted by $\mathcal{P}(f(\ln d_L)|\tau)$, which can be rewritten by marginalizing over t ,

$$\mathcal{P}(f(\ln d_L)|\tau) = \prod_{i=1}^{n_{\text{iter}}} \int dt_i \mathcal{P}(f(d_{L,i}), t|\tau) = \prod_{i=1}^{n_{\text{iter}}} \int dt_i \mathcal{P}(t_i) \mathcal{P}(f(\ln d_{L,i})|t, \tau). \quad (\text{A.3})$$

Here, the contraction factors t are beta distributed, i.e., $\mathcal{P}(t_i) = \text{Beta}(t_i|1, n_{\text{live}})$. The tricky part is to rewrite $\mathcal{P}(f(\ln d_{L,i})|t, \tau)$ such that we can integrate over t_i . Under the assumption that we can find a unique solution for $f(\ln d_{L,i})$ with $i = 1, \dots, n_{\text{iter}}$ for each of the following n_{iter} equations using eq. (A.2),

$$X(f(\ln d_{L,i})|\tau) = \prod_{j=1}^i t_j \quad \text{for } i = 1, \dots, n_{\text{iter}} \quad (\text{A.4})$$

the distribution $\mathcal{P}(f(\ln d_{L,i})|\tau, t)$ is defined via

$$\mathcal{P}(f(\ln d_{L,i})|\tau, t) = \delta\left(X(f(\ln d_{L,i})|\tau) - \prod_{j=1}^i t_j\right) \left| \frac{\partial X(f(\ln L)|\tau)}{\partial f(\ln L)} \right|_{f(\ln L)=f(\ln d_{L,i})}. \quad (\text{A.5})$$

In the following, we will denote $|\partial_\tau X|_{d_{L,i}} = \left| \frac{\partial X(f(\ln L)|\tau)}{\partial f(\ln L)} \right|_{f(\ln L)=f(\ln d_{L,i})}$ and $X_i = X(f(\ln d_{L,i})|\tau)$, for the matter of brevity of the equations. With the definition of $t_0 := 1$, we can further rewrite eq. (A.5) to,

$$\mathcal{P}(f(\ln d_{L,i})|\tau, t) = \delta\left(t_i - \frac{X_i}{\prod_{j=0}^{i-1} t_j}\right) |\partial_\tau X|_{d_{L,i}} \prod_{j=0}^{i-1} \frac{1}{t_j}, \quad (\text{A.6})$$

$$= \delta\left(t_i - \frac{X_i}{X_{i-1}}\right) |\partial_\tau X|_{d_{L,i}} \prod_{j=0}^{i-1} \frac{1}{t_j}, \quad (\text{A.7})$$

Using this and eq. (A.3), we can write down the overall reconstruction likelihood for this approach:

$$\mathcal{P}(f(\ln d_L)|\tau) = \prod_{i=1}^{n_{\text{iter}}} \int dt_i \mathcal{P}(t_i) \delta\left(t_i - \frac{X_i}{X_{i-1}}\right) |\partial_\tau X|_{d_{L,i}} \prod_{j=0}^{i-1} \frac{1}{t_j} \quad (\text{A.8})$$

$$= \prod_{i=1}^{n_{\text{iter}}} \text{Beta}\left(\frac{X_i}{X_{i-1}} | 1, n_{\text{live}}\right) \frac{1}{X_{i-1}} |\partial_\tau X|_{d_{L,i}} \quad (\text{A.9})$$

$$= \prod_{i=1}^{n_{\text{iter}}} n_{\text{live},i} \frac{X_i^{n_{\text{live},i}-1}}{X_{i-1}^{n_{\text{live},i}}} |\partial_\tau X|_{d_{L,i}} \quad (\text{A.10})$$

So far, the likelihood model and the prior model can be used for a maximum a posteriori approximation of $f(\ln L)$. For variational inference (VI), however, the Fisher metric is required,

$$(I_F)_{k,l} = \int df(\ln d_{L,0}) \dots \int df(\ln d_{L,n_{\text{iter}}}) \frac{\partial}{\partial \tau_k} (-\ln(\mathcal{P}(f(\ln d_L)|\tau))) \quad (\text{A.11})$$

$$\times \frac{\partial}{\partial \tau_l} (-\ln(\mathcal{P}(f(\ln d_L)|\tau))). \quad (\text{A.12})$$

Due to the complexity and length of the calculations involved in this integration and the difficulties of its implementation, we have opted to use the Gaussian approximation and reserve this approach for future work.

A.2 Reparametrization

The aim of the reparametrization is to map the likelihood–prior–volume function such that the correlated field describing the log-normal process is constant in the Gaussian case and

non-Gaussianity is modelled by structures in the correlated field (section 2.2.2). Below, we show that the reparametrization in eq. (2.18) leads to this condition. For the Gaussian case (eq. (2.23)), we have

$$\frac{d \ln L}{d \ln X} = \frac{d}{d \ln X} \left(-\frac{X^{2/C}}{2\sigma^2} \right) = \frac{2}{C} \ln L. \quad (\text{A.13})$$

For the Gaussian case, we know $\ln L_{\max} = 0$ analytically. If we substitute this into eq. (2.18) we find the desired condition,

$$\frac{df(\ln L)}{d \ln X} = -\frac{2}{C} = -e^{-\tau(\ln X)}, \quad (\text{A.14})$$

leading to $\tau(\ln X) = \text{const.}$ It can be seen that the corresponding reparameterization is given by the following equation

$$f(\ln L) = -\ln \left(\ln \frac{L_{\max}}{L} \right). \quad (\text{A.15})$$

A.3 Maximal likelihood calculation

We calculate the maximum likelihood given the data on the likelihood dead contours d_L . In Handley and Lemos (2019) the maximum of the Shannon entropy $\mathcal{I}_{\max} = \max \left[\ln \left(\frac{\mathcal{P}(\theta_M|d)}{\mathcal{P}(\theta_M)} \right) \right]$ is calculated via,

$$\mathcal{I}_{\max} = H + \frac{C}{2}, \quad (\text{A.16})$$

where C is the dimension of the problem. This gives the maximum log-likelihood $\ln L_{\max}$,

$$\ln \left(\frac{L_{\max}}{Z} \right) = \int d\theta_M \mathcal{P}(\theta_M|d) \ln \left(\frac{L}{Z} \right) + \frac{C}{2} \quad (\text{A.17})$$

$$\rightarrow \ln L_{\max} = \langle \ln L \rangle_{(\theta_M|d)} + \frac{C}{2}. \quad (\text{A.18})$$

This can be approximated using the data on the likelihood contours and taking the deterministic NS information on the prior volumes $\{d_{X,i}\}$ according to eq. (2.8),

$$\ln L_{\max} = \int d\theta \frac{\mathcal{P}(d|\theta)\mathcal{P}(\theta)}{Z} \ln L + \frac{C}{2} \quad (\text{A.19})$$

$$\approx \sum_{i=1}^{n_{\text{iter}}} \frac{1}{2} \frac{d_{L,i}(d_{X,i-1} - d_{X,i+1})}{Z_d} \ln d_{L,i} + \frac{C}{2}. \quad (\text{A.20})$$

Here, Z_d is the evidence calculated according to eq. (2.7), with $\{d_{L,i}\}$ as the likelihood information and $\{d_{X,i}\}$ as the prior volume information.

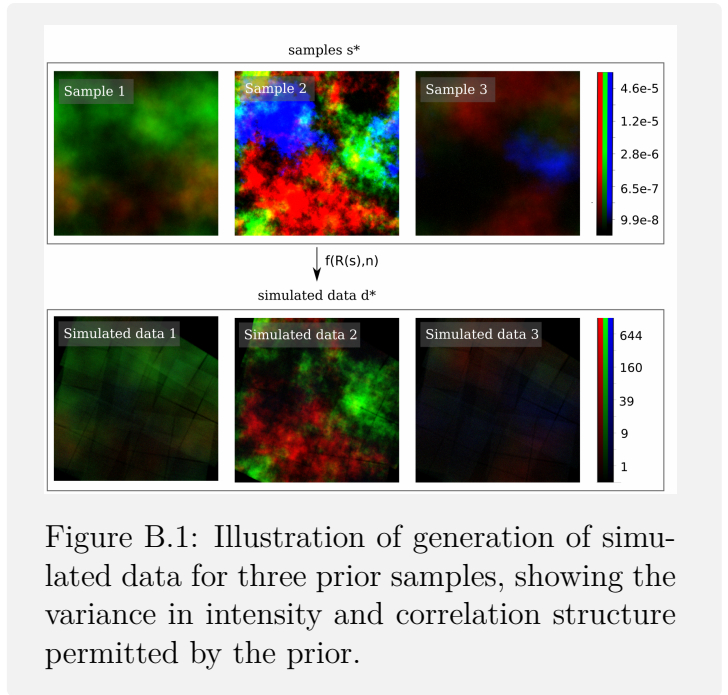
Appendix B

Supplementary material chapter 3

B.1 Synthetic data generation

Given the generative models, we can construct prior samples of the individual components and of the imaged sky composed of them, as described in section 3.5.

Three of these prior samples are shown at the top of fig. B.1. They illustrate how the prior samples are converted into simulated data using the instrument response and mimicking Poisson noise. As mentioned in section 3.5, we use the prior samples and the simulated data to fine-tune the hyper-parameters prior to reconstruction. As we can see by comparison, the chosen hyper-parameters ensure that the order of magnitude of the data in fig. 3.3 is the same as the order of magnitude of the simulated data. Moreover, and more importantly, the simulated data allow us to perform the validation of the algorithm, as described in section 3.7.



B.2 Computational analysis

In this section, we present a comparison of the introduced algorithm including the transition model and a pure MF reconstruction in terms of computational time and reconstruction

error. In case of the transition model, we started with a SF reconstruction and use the corresponding result as an initial condition for the MF reconstruction, as described in section 3.4. In the other case, we started the reconstruction on the whole MF parameter space from the beginning. We considered four different spatial resolutions, from 64×64 to 512×512 pixels, for which we generated simulated data and perform the corresponding reconstruction on a single core for the transition model and the pure MF model. This allowed us to compare, for each problem size, the time complexity at each iteration and the reconstruction error as a function of time.

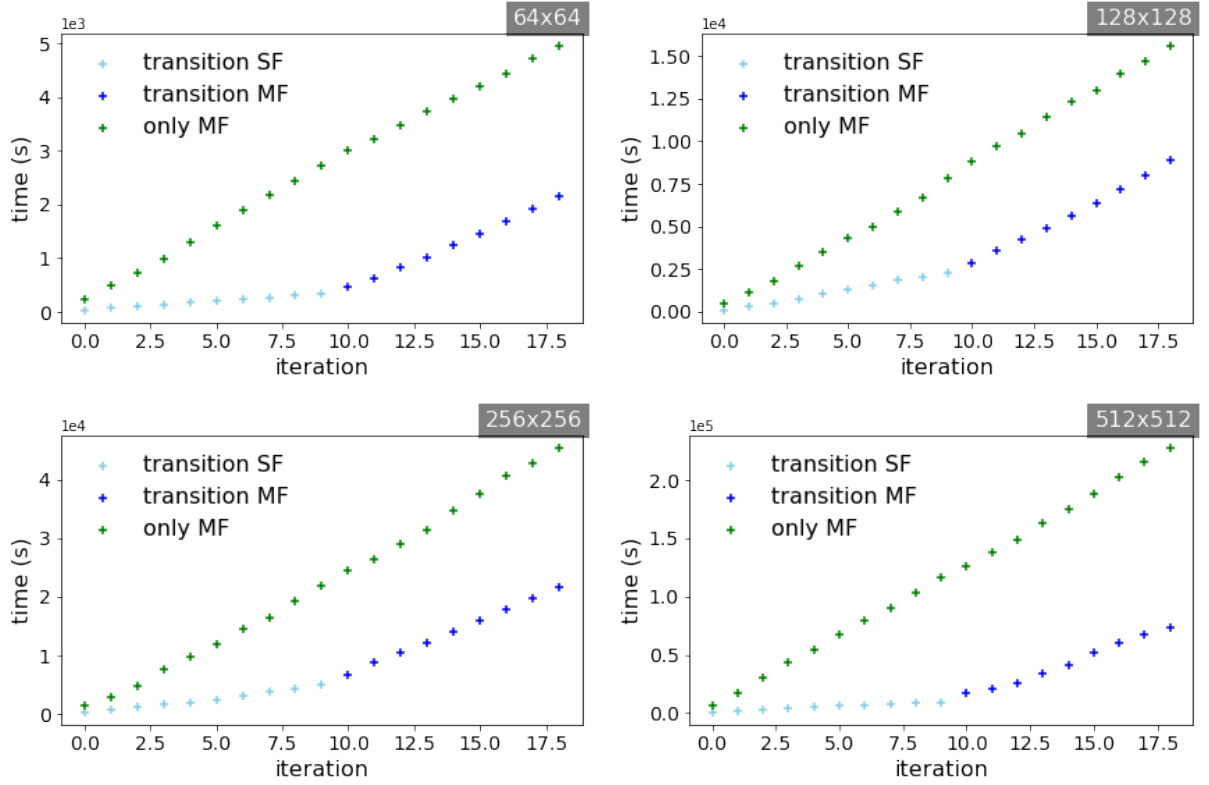


Figure B.2: Time complexity (top-left: 64×64 spatial pixels; top-right: 128×128 spatial pixels; bottom-left: 256×256 spatial pixels; bottom-right: 512×512 spatial pixels). The time complexity is plotted for the different models. In green, we show just the MF reconstruction times per iteration. Light blue: duration for each iteration in the SF model before the transition. Dark blue: duration of the MF model iterations after the transition. The first dark blue marker also includes the transition time.

As mentioned in section 3.4, the parameter space for the SF reconstruction is much smaller (table 3.2), leading to higher computational time for each iteration in the MF reconstruction. This effect is also shown in fig. B.2. The time complexity of each iteration is accordingly higher in the MF reconstruction than in the SF reconstruction, resulting in an overall lower time complexity for the transition model reconstruction. It can be seen that similarly to the duration of each iteration in the reconstruction, the transition time also

increases with the growing parameter space. However, the increased transition time is not significant compared to the overall time savings.

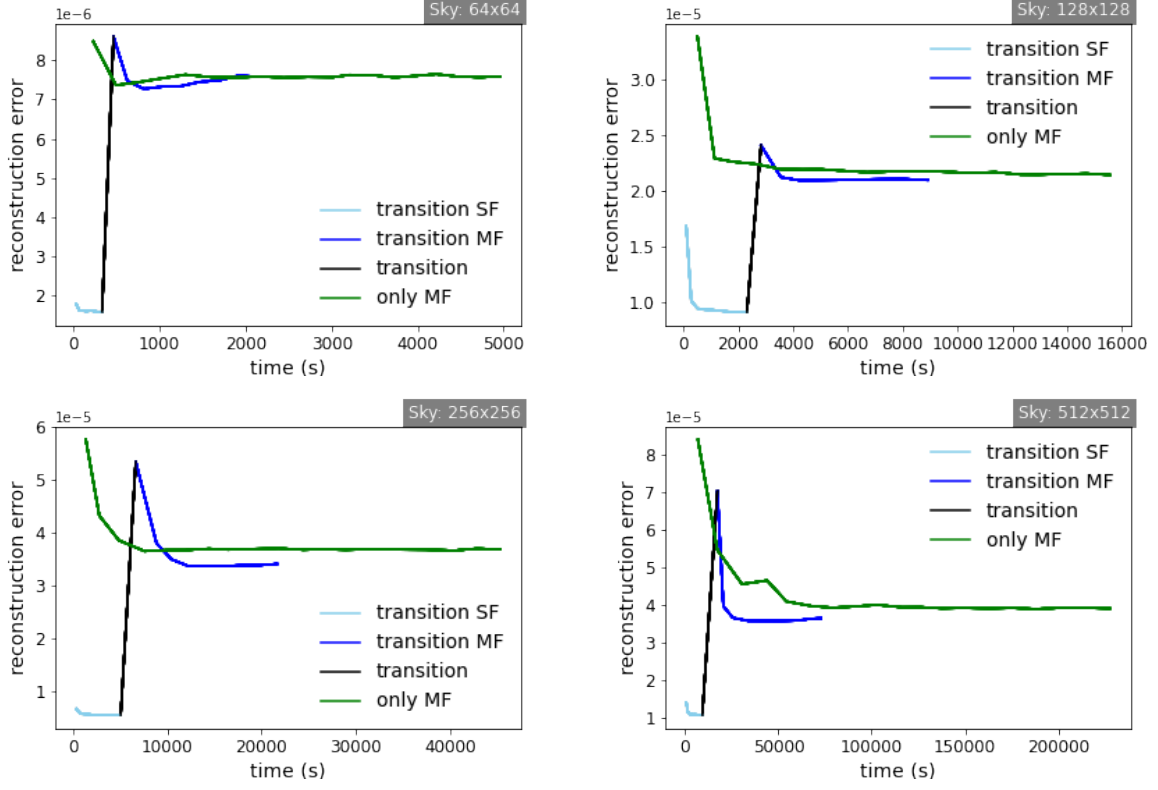


Figure B.3: Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged sky. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.

What is more important for the analysis is the notion of how the reconstruction error of the reconstruction behaves over time. This is shown for the different components in figs. B.3 to B.5. As a measure of the reconstruction error, we computed the posterior sample mean for N samples of the Frobenius norm of the sample residuals $\epsilon_r^* = (s^* - s_{\text{gt}})$ according to eq. (3.20) for each component:

$$\langle \|\epsilon_r\|_F \rangle_{s^*} = \frac{1}{N} \sum_{n=0}^N \left(\sum_{i,j,k} (\epsilon_{r,n}^*)_{i,j,k}^2 \right)^{\frac{1}{2}}, \quad (\text{B.1})$$

where i, j, k are the corresponding spatial and spectral pixel indices. Due to the consideration of a smaller space in the iterations of the SF model, we would typically expect a smaller reconstruction error in terms of small Frobenius norm. It can be seen that the computational advantage of the transition model approach increases as the problem size increases in terms of a higher number of spatial pixels. This is especially true for the diffuse component, which is constructed from an outer product of correlated fields.

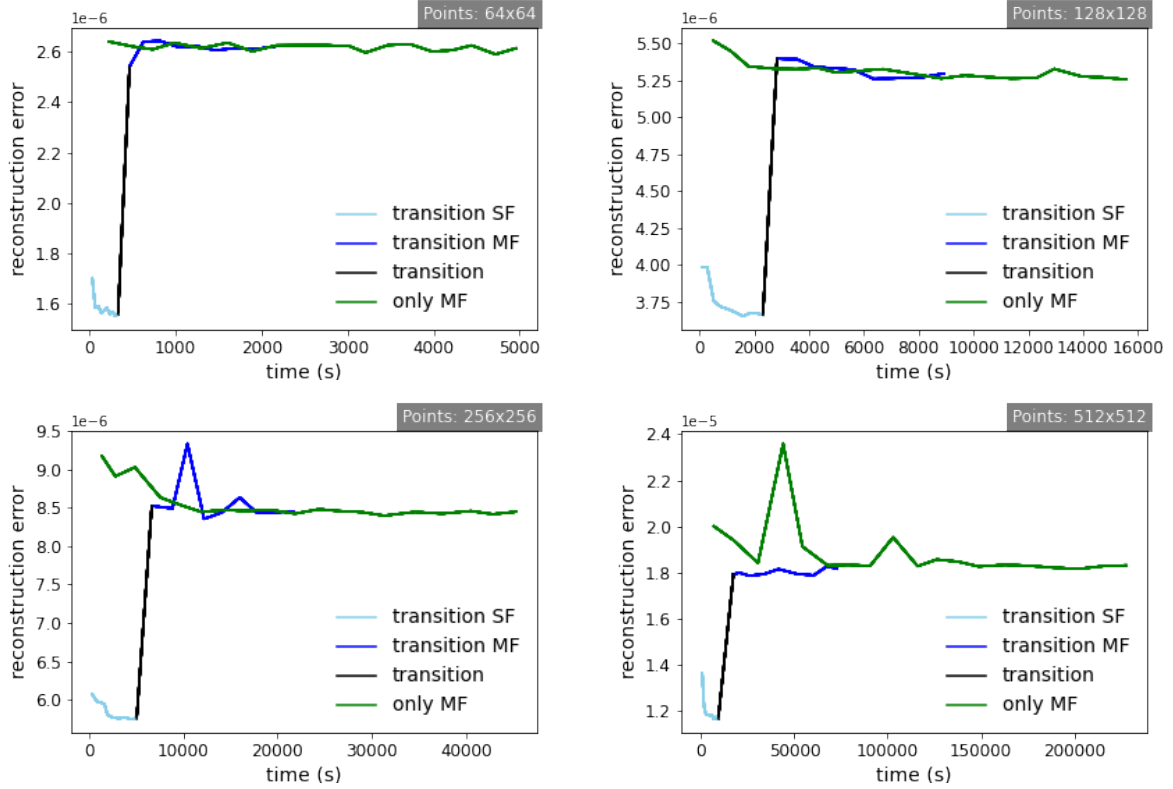


Figure B.4: Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged point source component. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.

B.3 Further diagnostics for synthetic data reconstruction

In this section, we present further diagnostic plots for sanity checks on the simulated data reconstruction in section 3.7. The analysis of these plots can be found in the according sections. We show the UWRs and residuals for the simulated data case in fig. B.6. Figure B.7 shows the reconstruction results for the simulated data case for each energy bin together with the associated uncertainty.

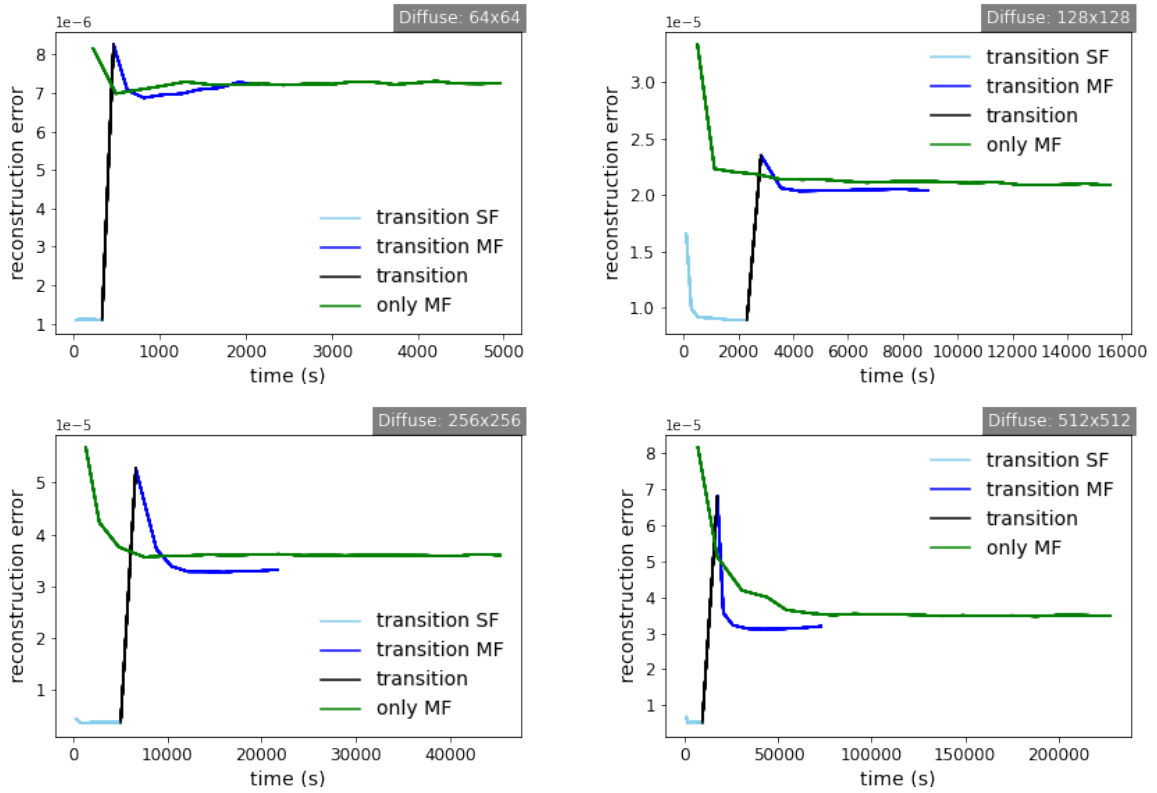


Figure B.5: Reconstruction error in terms of the Frobenius norm (eq. (B.1)) for 64×64 , 128×128 , 256×256 , and 512×512 spatial pixels for the imaged diffuse component. The green line marks the reconstruction error as a function of time for the pure MF reconstruction. The light blue line marks the reconstruction error of the SF reconstruction as part of the transition model and, correspondingly, the black line marks the transition and the blue line marks the subsequent transition model MF reconstruction. In the iterations of the SF model, we typically anticipate lower reconstruction error in terms of small Frobenius norm. This expectation is attributed to the model's consideration of a smaller space.

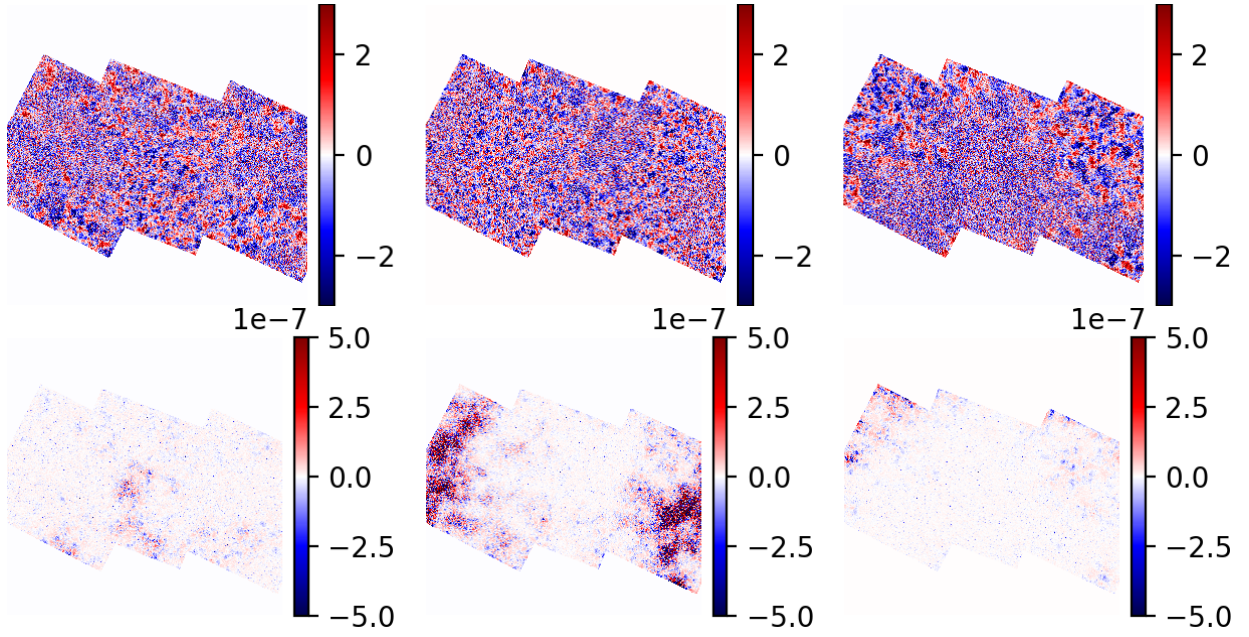


Figure B.6: Synthetic data reconstruction UWRs (top row) and residuals (bottom row) for the individual energy bins (left: 0.5-1.2 keV, center: 1.2-2.0 keV, right: 2.0-7.0 keV) according to eq. (3.19).

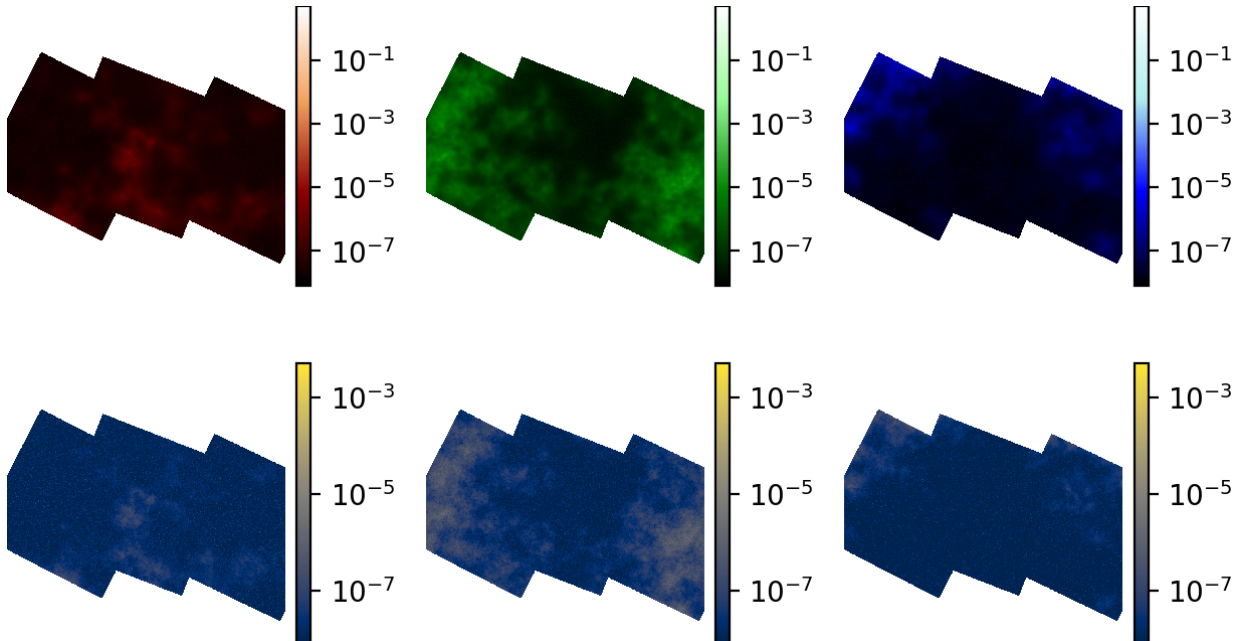


Figure B.7: Synthetic data reconstruction uncertainties for the individual energy bins (left: 0.5-1.2 keV, center: 1.2-2.0 keV, right: 2.0-7.0 keV). Top row: Reconstruction results for the flux in $[s^{-1} cm^{-2}]$ for the individual energy bins. Bottom row: Uncertainty maps for the individual energy bins.

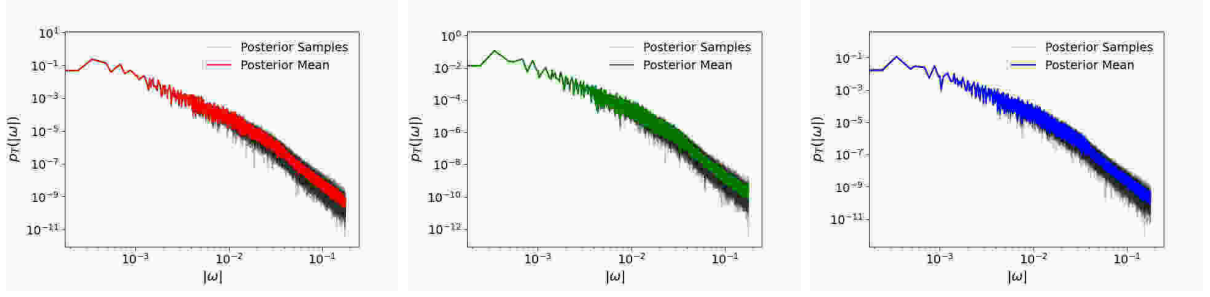


Figure B.8: Spatial power spectra of the posterior mean and samples in the diffuse component for each energy bin (red: 0.5-1.2 keV, green: 1.2-2.0 keV, blue: 2.0-7.0 keV)

B.4 Further diagnostics for SN1006 reconstruction

Here, we show more diagnostic plots for the analysis of the reconstruction results presented in section 4.4. First, the reconstruction mean and posterior samples of the spatial power spectrum are shown in fig. B.8. For further analysis of the reconstruction of the sky flux of the remnant SN1006, we present the posterior standard deviation separately for each energy bin and accompanied by the corresponding color bars in fig. B.9. Finally, table B.1 shows the NWRs according to eq. (3.23) for each energy bin and dataset.

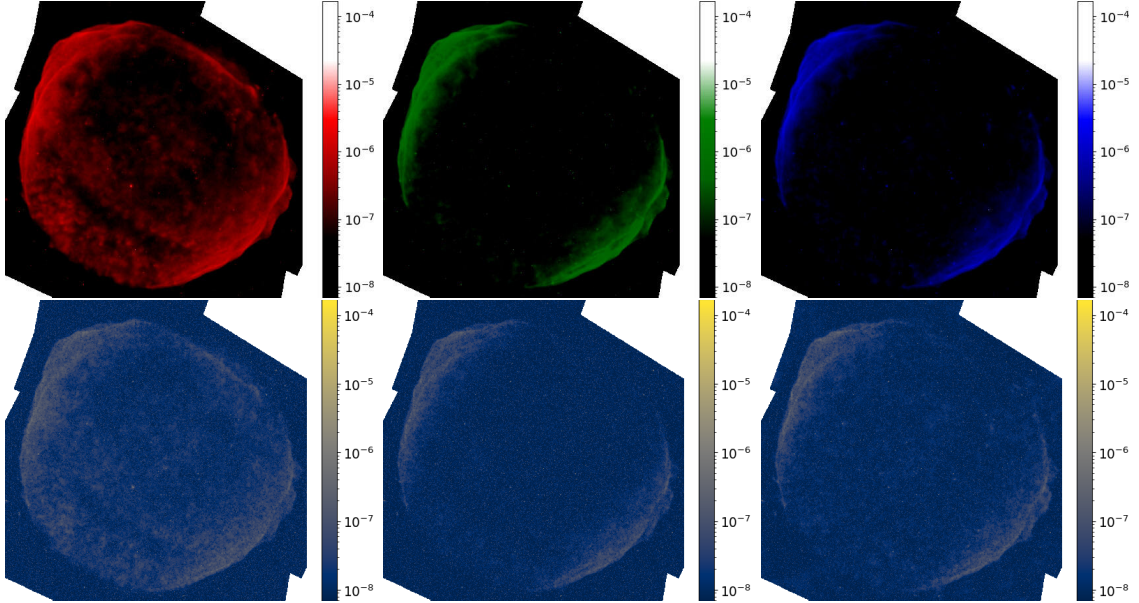
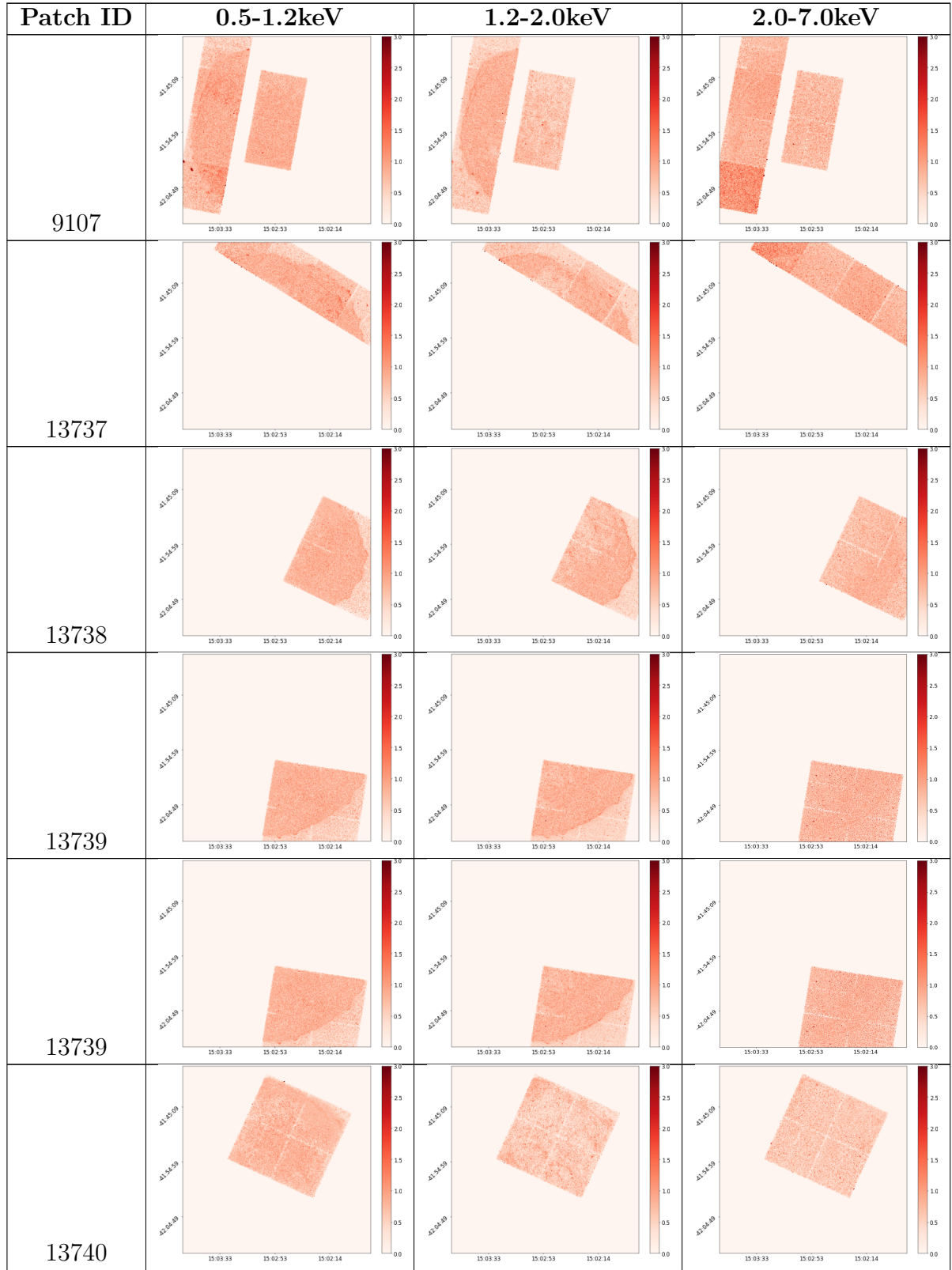


Figure B.9: Posterior means and standard deviations for each energy bin in $[s^{-1} \text{ cm}^{-2}]$: Top row: Posterior means (red: 0.5-1.2 keV, green: 1.2-2.0 keV, blue: 2.0-7.0 keV). Bottom row: Posterior standard deviations (left: 0.5-1.2 keV, center: 1.2-2.0 keV, right: 2.0-7.0 keV).



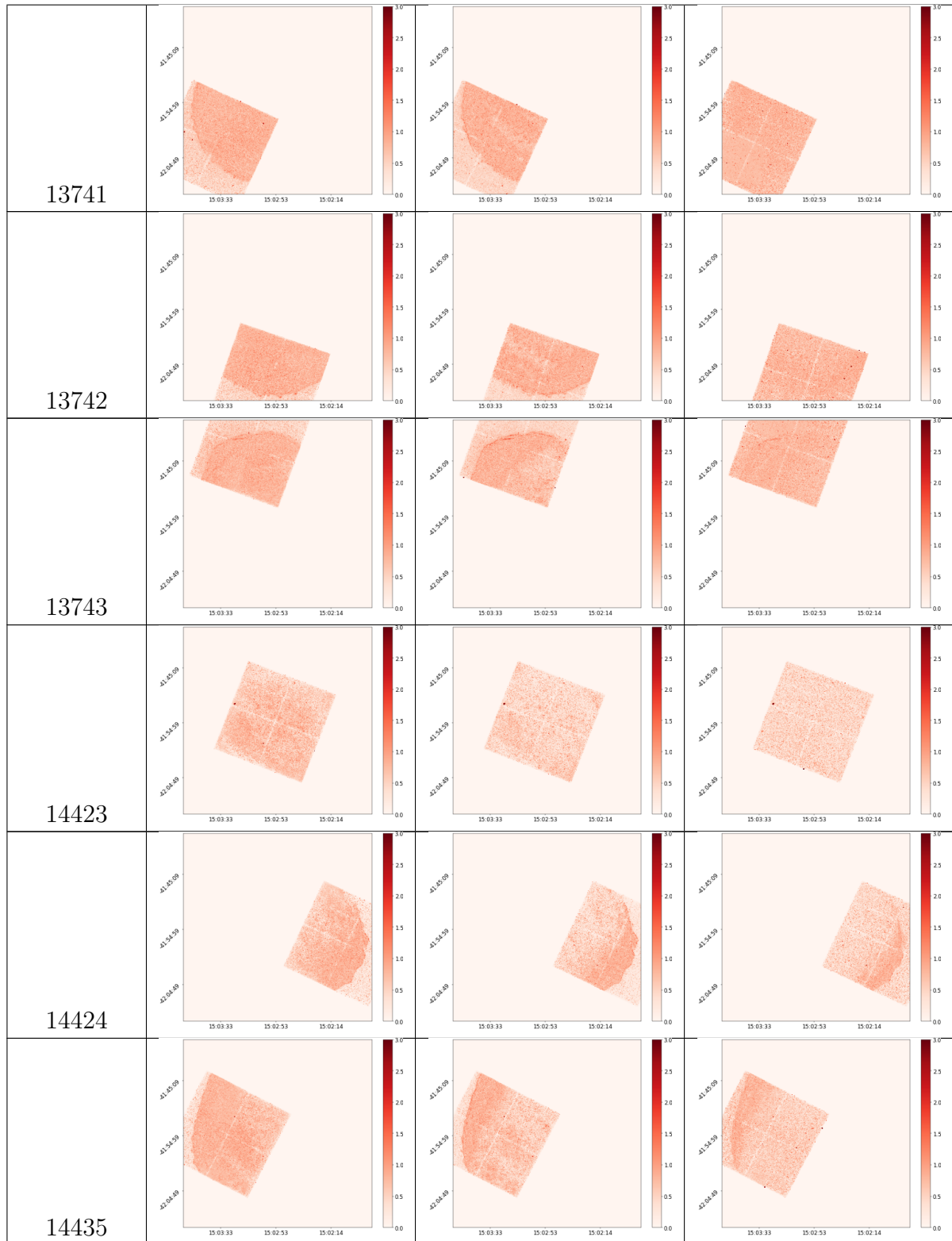


Table B.1: NWRs (eq. (3.23)) for each dataset in table 3.3 and energy bin.

Appendix C

Supplementary material chapter 4

C.1 eROSITA data: LMC SN1987A

First draft written by Margret Westerkamp.

The LMC SN1987A EDR data were pre-processed using the eSASS pipeline, which is described in detail by Brunner et al. (2018) and Predehl et al. (2021)^a. In particular, the data was extracted and manipulated using the eSASS `evtool` command. We list the flag values we chose for the `evtool` command^b in table C.1. We computed the exposure maps for the eROSITA event files using the eSASS `expmap` command and the corresponding flags in table C.2. The data per energy bin and per TM is shown in fig. C.1. The corresponding exposure maps summed over the 5 TMs are shown in appendix C. The PSFs used for the PSF linear patched convolution representation can be found in `tm[1-7]_2dpsf_190219v05.fits` in the CALDB. The effective area for the individual CA can be found in `tm[1-7]_arf_filter_000101v02.fits` in the CALDB.

^aFurther information on the eSASS pipeline can also be found at <https://erosita.mpe.mpg.de/edr/DataAnalysis/>.

^bA further description of the flags can be found at https://erosita.mpe.mpg.de/edr/DataAnalysis/evtool_doc.html.

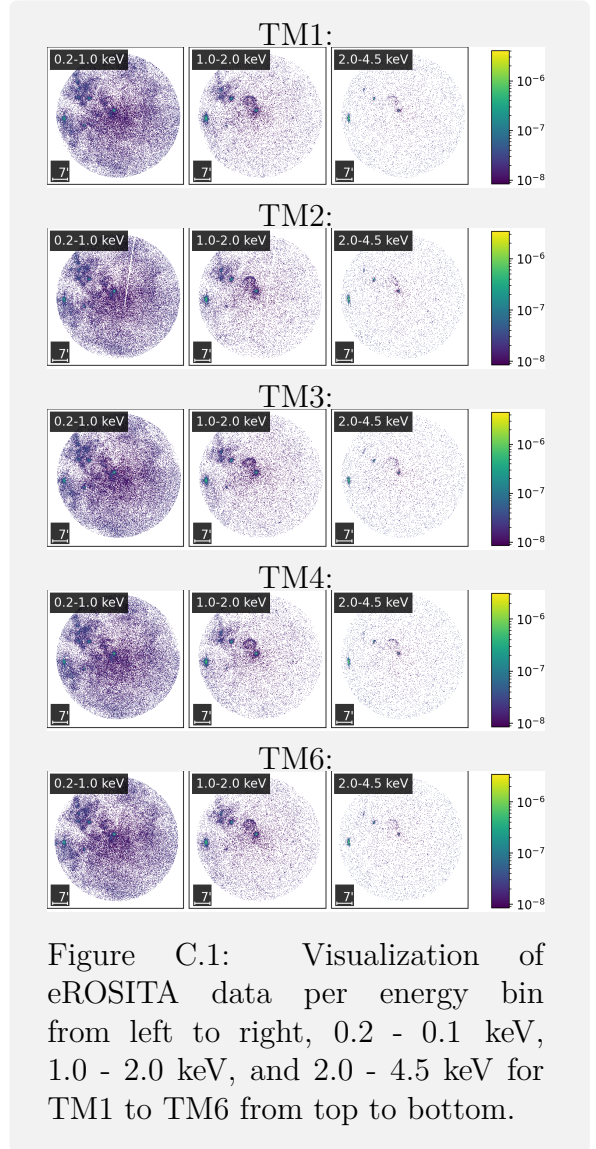


Table C.1: Flags and their corresponding data types for `evtool`, where `tmid` $\in \{1, 2, 3, 4, 6\}$ and $(\text{emin}, \text{emax}) \in \{(0.2, 1.0), (1.0, 2.0), (2.0, 4.5)\}$

Flag	Data Type	Value
<code>clobber</code>	bool	True
<code>events</code>	bool	True
<code>image</code>	bool	True
<code>size</code>	int	1024
<code>rebin</code>	int	80
<code>center_position</code>	tuple	None
<code>region</code>	str	None
<code>gti</code>	str	None
<code>flag</code>	str	None
<code>flag_invert</code>	bool	None
<code>pattern</code>	int	15
<code>telid</code>	int	tmid
<code>emin</code>	float str	emin
<code>emax</code>	float str	emax
<code>rawxy</code>	str	None
<code>rawxy_telid</code>	int	None
<code>rawxy_invert</code>	bool	False
<code>memset</code>	int	None
<code>overlap</code>	float	None
<code>skyfield</code>	str	None

Table C.2: Flags and their corresponding data types for `expmap`, where `tmid` $\in \{1, 2, 3, 4, 6\}$ and $(\text{emin}, \text{emax}) \in \{(0.2, 1.0), (1.0, 2.0), (2.0, 4.5)\}$

Parameter	Data Type	Value
<code>emin</code>	float str	emin
<code>emax</code>	float str	emax
<code>withsinglemaps</code>	bool	True
<code>withmergedmaps</code>	bool	False
<code>gtitype</code>	str	GTI
<code>withvignetting</code>	bool	True
<code>withdetmaps</code>	bool	True
<code>withweights</code>	bool	True
<code>withfilebadpix</code>	bool	True
<code>withcalbadpix</code>	bool	True
<code>withinputmaps</code>	bool	False

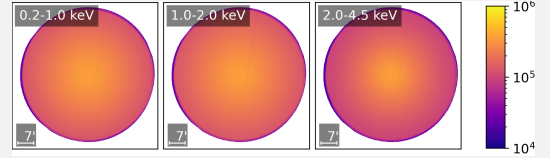


Figure C.2: eROSITA exposure maps summed over all 5 TMs.

C.2 Hyper-parameter search

First draft written by Margret Westerkamp.

The prior model described in section 4.3.2 requires choosing a set of hyper-parameters, which describe the mean and standard deviation of the Gaussian processes modelling the prior. The meaning of the specific hyper-parameters of the correlated field is more precisely described in Arras et al. (2022). In particular, the offset mean of the correlated field parametrizes the mean of τ in eq. (4.3) and therefore the mean of the log-normal flux. Accordingly, we take the exposure-corrected data, d_e , shown in fig. 4.1 and calculate its mean $\langle d_e \rangle$ and set both the offset mean of φ_{\ln} and $\varphi_{\ln,b}$ to $\log \langle d_e \rangle = -19.9$.

We use the information on the detection threshold to set the hyper-parameters for the inverse gamma distribution used for the point sources. In particular we set the mean of the inverse gamma distribution as the sum of all fluxes from point sources which are higher than the detection threshold θ divided by the total number of pixels.

To determine the detection threshold θ , we set a minimum S/N, γ_{\min} , that is required to

reliably detect a source. Essentially, for a source to be detected, the S/N, γ , in each pixel must be higher than this set threshold, γ_{\min} . Specifically, for Poisson data, the S/N is given by $\gamma = \sqrt{\lambda}$, where λ is the expected number of counts in a pixel. We set γ_{\min} based on the confidence level we want for detection. In this case, we aim for a 99% confidence level, meaning there is a 99% probability that any observed signal is not just a random fluctuation,

$$\mathcal{P}(k \geq 1|\lambda) = 1 - \mathcal{P}(k = 0|\lambda) = 1 - e^{-\lambda} \stackrel{!}{=} 0.99, \quad (\text{C.1})$$

which leads to $\lambda_{\min} = 4.6$ and, consequently, $\gamma_{\min} = \sqrt{\lambda_{\min}} = 2.14$. The pixel-wise detection threshold θ_i is then defined, via the smallest flux, which can be reliably detected in each pixel i , which is given via λ_{\min} and the exposure in the corresponding pixel, E_i

$$\theta_i = \frac{\lambda_{\min}}{E_i}. \quad (\text{C.2})$$

We want to find an overall detection threshold for the whole image, which is then defined via maximal exposure, E_{\max}

$$\theta = \frac{\lambda_{\min}}{E_{\max}} = 2.5 \times 10^{-9}. \quad (\text{C.3})$$

Eventually, this leads to a mean $\mu_{\text{ig}} = 2.08 \times 10^{-9}$ of the inverse gamma distribution. The mode, Mo_{ig} , of the inverse gamma distribution should be even further below the detection threshold. In particular, we thus assume that the S/N for the mode is much lower, i.e. $\gamma_{\min} = 0.1$,

$$\text{Mo}_{\text{ig}} = \frac{0.1^2}{E_{\max}}. \quad (\text{C.4})$$

Having the mean and the mode, we can use these in order to calculate the hyper-parameters κ and β of the inverse gamma distribution via

$$\kappa = \frac{2}{\frac{\mu_{\text{ig}}}{\text{Mo}_{\text{ig}}} - 1} + 1, \quad (\text{C.5})$$

$$\beta = \text{Mo}_{\text{ig}}(\kappa + 1). \quad (\text{C.6})$$

A prior sample drawn from the prior given these hyper-parameters can be seen in fig. 4.2.

C.3 Results diagnostics

First draft written by Margret Westerkamp.

Here, we show additional plots corresponding to the reconstruction of SN1987A in the LMC as seen by SRG/eROSITA in the CalPV phase. First, we display the reconstruction shown in fig. 4.4 per energy bin in fig. C.5 to give a better understanding of the color bar used. We also show the corresponding uncertainty in the form of the standard deviation per energy bin in fig. C.6. Finally, we also performed single TM reconstructions, the results of which are shown per TM in fig. C.3. Important diagnostic measures to check for possible calibration inconsistencies in the single TM reconstructions are the NWRs (eq. (4.20)), which are shown per TM and energy bin in fig. C.8.

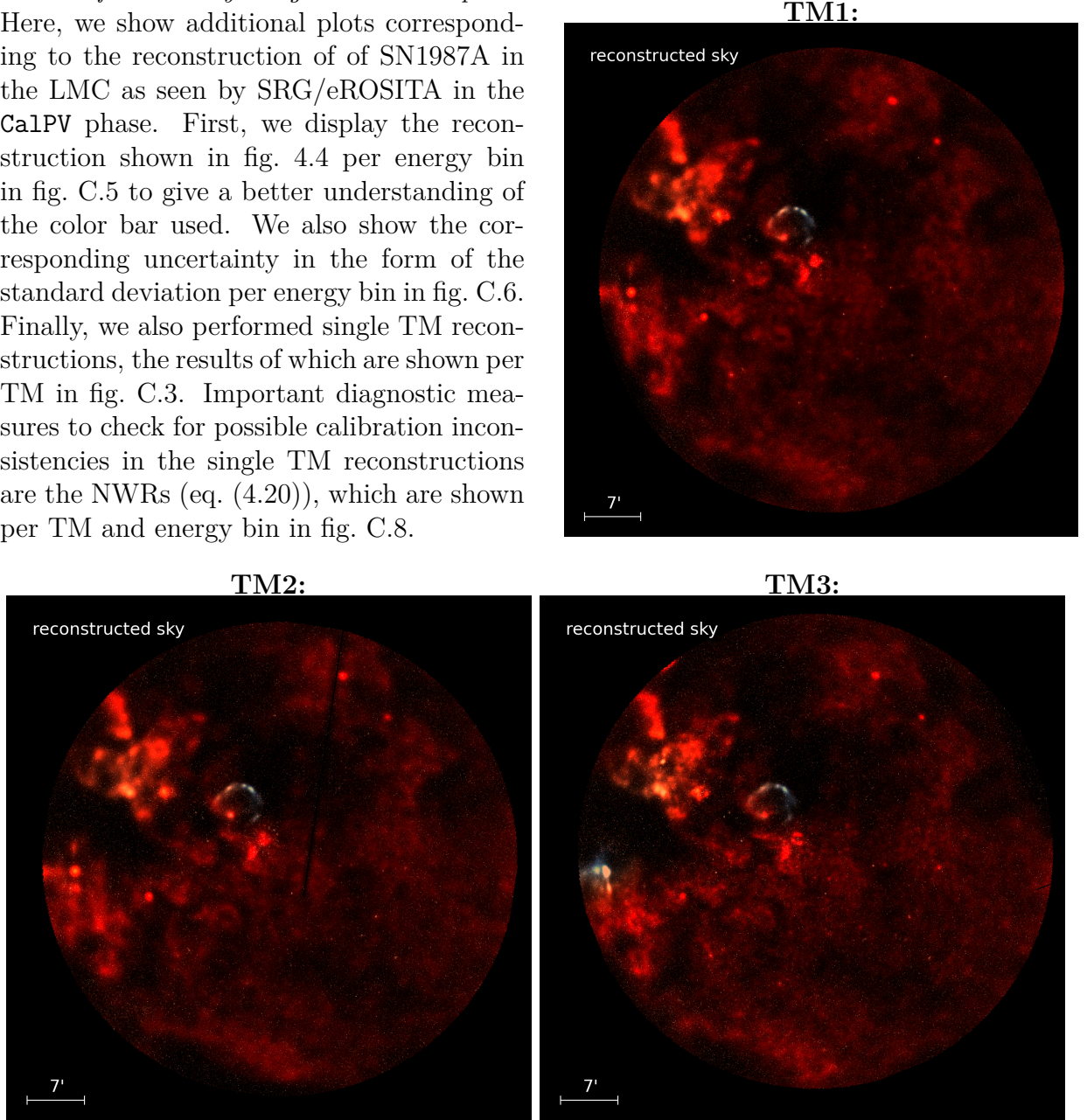


Figure C.3: Results for single-TM reconstructions for TM1, TM2 and TM3. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively. Continued for TM4 and TM6 in fig. C.4

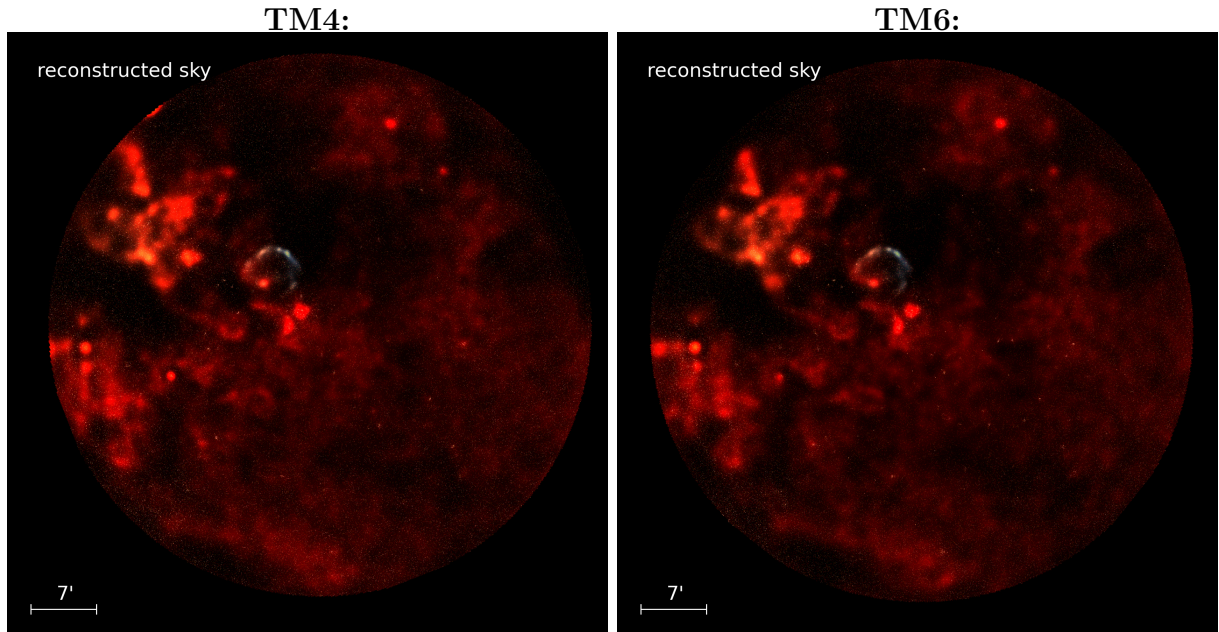


Figure C.4: Results for single-TM reconstructions for TM4 and TM6. The different colors represent the intensities in the three energy channels 0.2 - 0.1 keV, 1.0 - 2.0 keV, and 2.0 - 4.5 keV and are depicted in red, green, and blue, respectively.

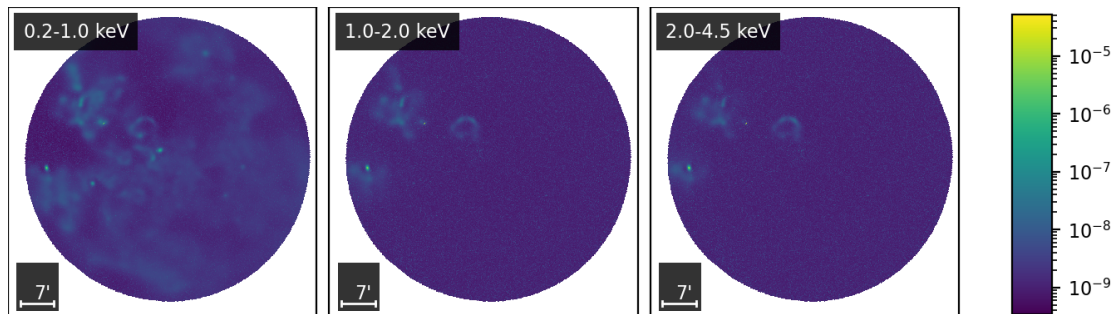


Figure C.5: Posterior mean of the sky reconstruction in the different energy bins.

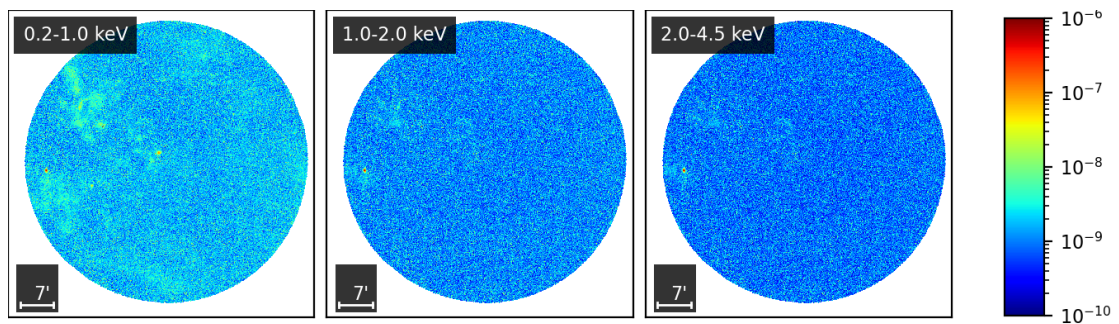


Figure C.6: Standard deviation per energy bin for the reconstruction shown in fig. 4.4 and fig. C.5.

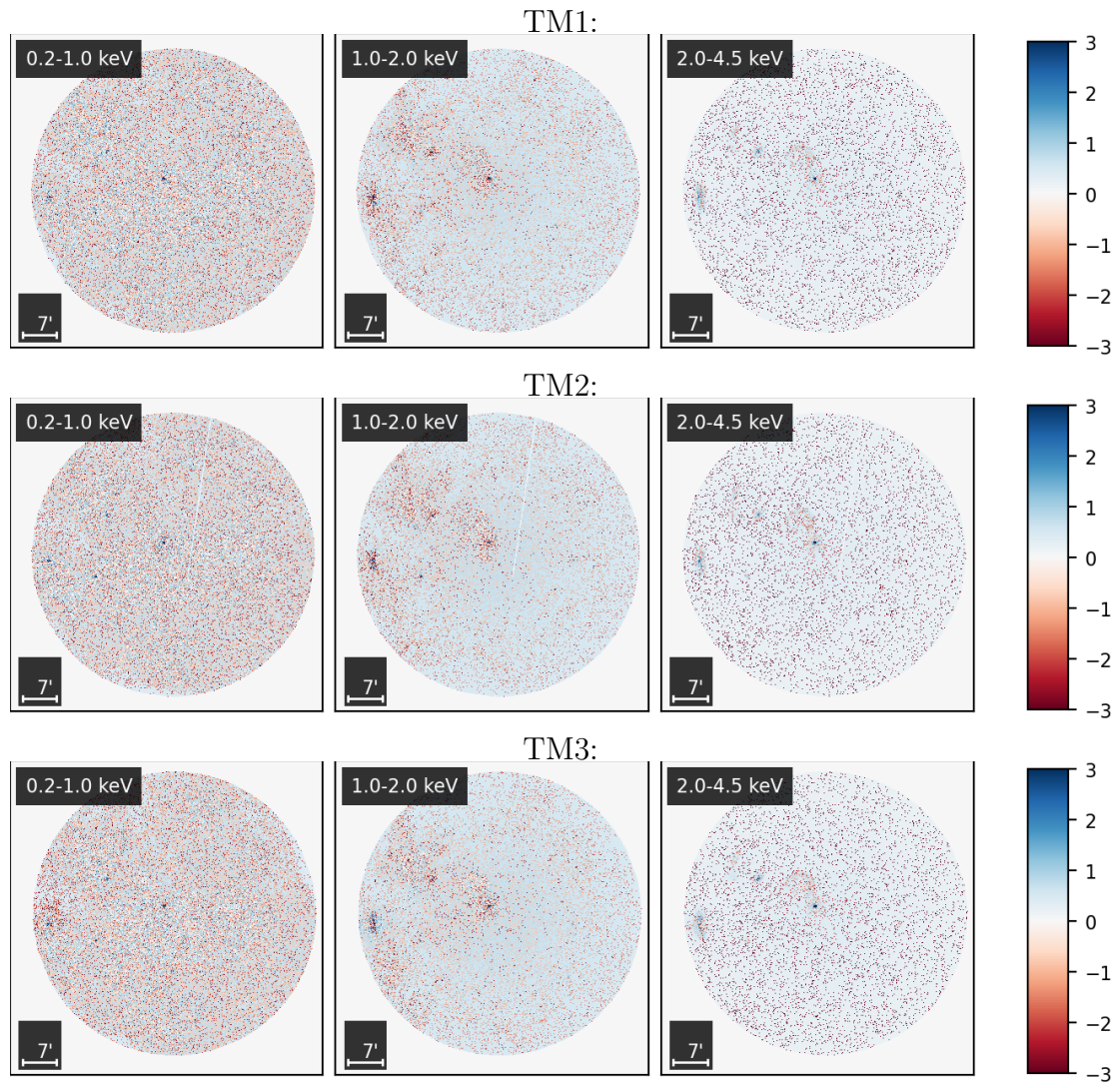


Figure C.7: Continued in fig. C.8.

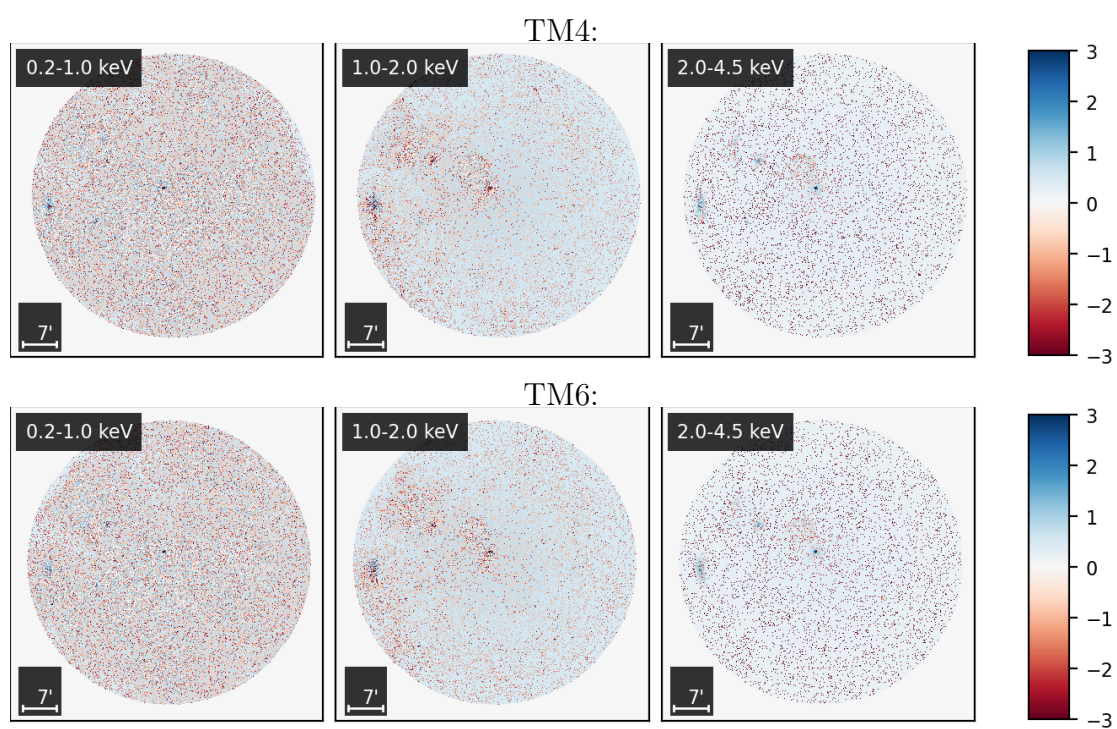


Figure C.8: Posterior mean of the NWRs for single-TM reconstructions.

C.4 Validation diagnostics

First draft written by Margret Westerkamp.

This section provides supplementary plots that offer further insights into the validation analysis discussed in section 4.5.1. These plots show in particular the images of the simulated sky, the simulated data, and the corresponding reconstruction. These are shown as an RGB image in fig. 4.5 as plots per energy, which serve to enhance the understanding of the color bar for the RGB image. In particular the data per energy bin and TM is presented in fig. C.9, the underlying simulated sky per energy bin is shown in fig. C.10 and the reconstructed sky per energy bin is illustrated in fig. C.11. Furthermore, we display the uncertainty in the reconstruction by means of the standard deviation for each energy bin in fig. C.12. Figure C.13 shows the standardized error for each energy bin.

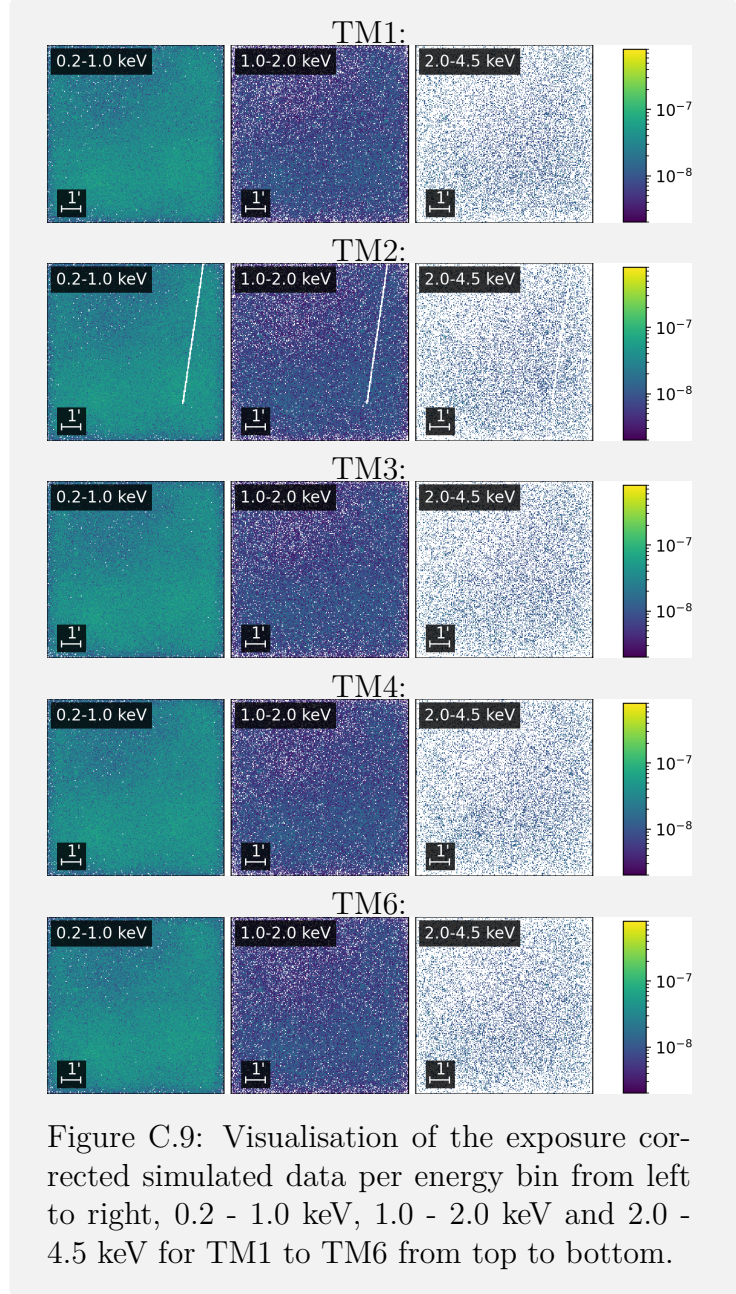


Figure C.9: Visualisation of the exposure corrected simulated data per energy bin from left to right, 0.2 - 1.0 keV, 1.0 - 2.0 keV and 2.0 - 4.5 keV for TM1 to TM6 from top to bottom.

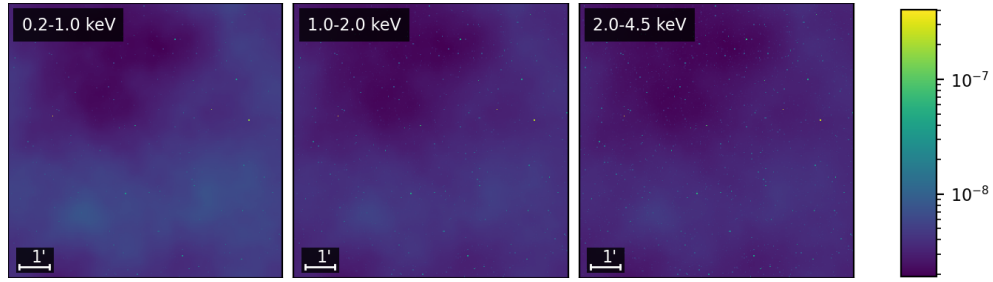


Figure C.10: Visualization of the simulated sky per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).

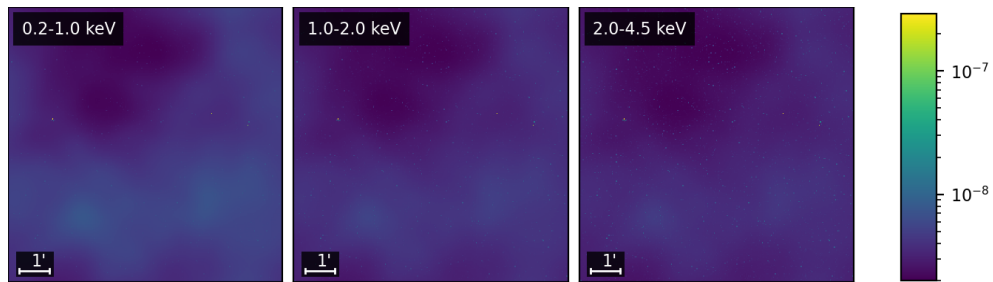


Figure C.11: Visualization of the reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).

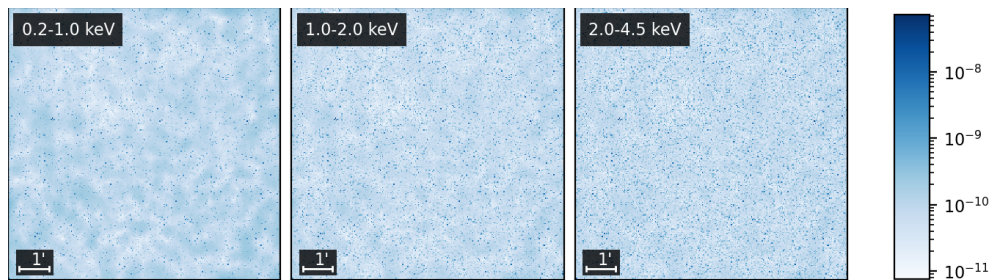


Figure C.12: Visualization of the standard deviation of the validation reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV)

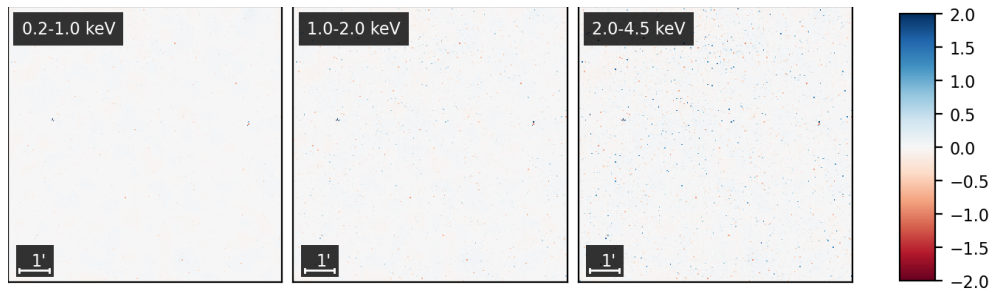


Figure C.13: Visualization of the standardized error of the validation reconstruction per energy bin (left: 0.2-1.0 keV, center: 1.0-2.0 keV, right: 2.0-4.5 keV).

Bibliography

- D. M. Alexander, F. E. Bauer, W. N. Brandt, D. P. Schneider, A. E. Hornschemeier, C. Vignali, A. J. Barger, P. S. Broos, L. L. Cowie, G. P. Garmire, L. K. Townsley, M. W. Bautz, G. Chartas, and W. L. W. Sargent. The Chandra Deep Field North Survey. XIII. 2 Ms Point-Source Catalogs. *The Astronomical Journal*, 126(2):539, 2003. doi: 10.1086/376473.
- K. Arnaud, R. Smith, and A. Siemiginowska. *Handbook of X-ray Astronomy*. Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press, 2011.
- P. Arras, M. Baltac, T. A. Enßlin, P. Frank, S. Hutschenreuter, J. Knollmueller, R. Leike, M. Newrzella, L. Platz, M. Reinecke, et al. Nifty5: Numerical information field theory v5. *Astrophysics Source Code Library*, pages ascl–1903, 2019.
- P. Arras, P. Frank, P. Haim, J. Knollmüller, R. Leike, M. Reinecke, and T. Enßlin. Variable structures in M87* from space, time and frequency resolved interferometry. *Nature Astronomy*, 6(2):259–269, 2022. doi: 10.1038/s41550-021-01548-0.
- G. Ashton, N. Bernstein, J. Buchner, X. Chen, Gábor G. Csányi, A. Fowlie, F. Feroz, M. Griffiths, W. Handley, M. Habeck, E. Higson, M. Hobson, A. Lasenby, D. Parkinson, L. B. Pártay, M. Pitkin, D. Schneider, J. S. Speagle, L. South, J. Veitch, P. Wacker, D. J. Wales, and D. Yallup. Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2(1), 2022. doi: 10.1038/s43586-022-00121-x.
- A. Bamba, R. Yamazaki, M. Ueno, and K. Koyama. Small-Scale Structure of the SN 1006 Shock with Chandra Observations. *The Astrophysical Journal*, 589(2):827–837, 2003. doi: 10.1086/374687.
- M. Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. 2017. doi: 10.48550/arXiv.1701.02434.
- D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017. doi: 10.1080/01621459.2017.1285773.
- J. Bobin, F. Sureau, and J.-L. Starck. Cosmic microwave background reconstruction from WMAP and Planck PR2 data. *Astronomy & Astrophysics*, 591:A50, 2016. doi: 10.1051/0004-6361/201527822.

- J. Bobin, I. El Hamzaoui, A. Picquenot, and F. Acero. Sparse BSS From Poisson Measurements. *IEEE Transactions on Image Processing*, 29:9429–9444, 2020. doi: 10.1109/TIP.2020.3027986.
- P. S. Broos, L. K. Townsley, E. D. Feigelson, K. V. Getman, F. E. Bauer, and G. P. Garmire. Innovations in the Analysis of Chandra-ACIS Observations. *The Astrophysical Journal*, 714(2):1582–1605, 2010. doi: 10.1088/0004-637x/714/2/1582.
- H. Brunner, T. Boller, D. Coutinho, T. Dauser, K. Dennerl, T. Dwelly, M. Freyberg, M. Fürmetz, A. Georgakakis, C. Grossberger, I. Kreykenbohm, G. Lamer, N. Meidinger, S. Müller, P. Predehl, J. Robrade, J. Sanders, and J. Wilms. eROSITA ground operations. In *Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray*, volume 10699 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 106995G, 2018. doi: 10.1117/12.2315138.
- H. Brunner, T. Liu, G. Lamer, A. Georgakakis, A. Merloni, M. Brusa, E. Bulbul, K. Dennerl, S. Friedrich, A. Liu, C. Maitra, K. Nandra, M. E. Ramos-Ceja, J. S. Sanders, I. M. Stewart, T. Boller, J. Buchner, N. Clerc, J. Comparat, T. Dwelly, D. Eckert, A. Finoguenov, M. Freyberg, V. Ghirardini, A. Gueguen, F. Haberl, I. Kreykenbohm, M. Krumpe, S. Osterhage, F. Pacaud, P. Predehl, T. H. Reiprich, J. Robrade, M. Salvato, A. Santangelo, T. Schrabback, A. Schwoppe, and J. Wilms. The eROSITA Final Equatorial Depth Survey eFEDS: X-ray catalogue. *Astronomy & Astrophysics*, 661:A1, 2022. doi: 10.1051/0004-6361/202141266.
- J. Buchner. Nested sampling methods. *Statistics Surveys*, 17:169 – 215, 2023. doi: 10.1214/23-SS144.
- J. Buchner and P. Boorman. *Statistical Aspects of X-ray Spectral Analysis*, pages 5403–5451. Springer Nature Singapore, 2024. ISBN 978-981-19-6960-7. doi: 10.1007/978-981-19-6960-7_175.
- D. N. Burrows, E. Michael, U. Hwang, R. McCray, R. A. Chevalier, R. Petre, G. P. Garmire, S. S. Holt, and J. A. Nousek. The X-Ray Remnant of SN 1987A. *The Astrophysical Journal*, 543(2):L149–L152, 2000. doi: 10.1086/317271.
- T. Calderwood, A. Dobrzycki, H. Jessop, and D. E. Harris. The Sliding-Cell Detection Program for Chandra X-ray Data. In *Astronomical Data Analysis Software and Systems X*, volume 238, page 443, 2001.
- G. Cassam-Chenaï, J. P. Hughes, E. M. Reynoso, C. Badenes, and D. Moffett. Morphological Evidence for Azimuthal Variations of the Cosmic-Ray Ion Acceleration at the Blast Wave of SN 1006. *The Astrophysical Journal*, 680(2):1180–1197, 2008. doi: 10.1086/588015.
- Chandra X-ray Center. *The Chandra Proposers’ Observatory Guide*, 2021. URL <https://cxc.harvard.edu/proposer/POG/index.html>. Version 23.0.

- N. Chopin and C. P. Robert. Properties of nested sampling. *Biometrika*, 97(3):741–755, 2010. doi: 10.1093/biomet/asq021.
- J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965. doi: 10.2307/2003354.
- R. G. Cruddace, G. H. Hasinger, and J. H. Schmitt. The application of a maximum likelihood analysis to detection of sources in the ROSAT data base. In *European Southern Observatory Conference and Workshop Proceedings*, volume 28, pages 177–182, 1988.
- J. Davis, M. Bautz, D. Dewey, R. Heilmann, J. Houck, D. Huenemoerder, H. Marshall, M. Nowak, M. Schattenburg, N. Schulz, and R. Smith. Raytracing with MARX: X-ray observatory design, calibration, and support. *Proceedings of SPIE: The International Society for Optical Engineering*, 8443, 2012. doi: 10.1117/12.926937.
- J. E. Davis. Event Pileup in Charge-coupled Devices. *The Astrophysical Journal*, 562(1): 575, 2001. doi: 10.1086/323488.
- L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986. ISBN 0387963057.
- K. K. Dyer, S. P. Reynolds, and K. J. Borkowski. Separating Thermal and Nonthermal X-Rays in Supernova Remnants. II. Spatially Resolved Fits to SN 1006 AD. *The Astrophysical Journal*, 600:752 – 768, 2003. doi: 10.1086/380093.
- H. Ebeling and G. Wiedenmann. Detecting structure in two dimensions combining Voronoi tessellation and percolation. *Physical Review E*, 47:704–710, 1993. doi: 10.1103/PhysRevE.47.704.
- V. Eberle, P. Frank, J. Stadler, S. Streit, and T. Enßlin. Butterfly Transforms for Efficient Representation of Spatially Variant Point Spread Functions in Bayesian Imaging. *Entropy*, 25(4):652, 2023. doi: 10.3390/e25040652.
- V. Eberle, M. Guardiani, M. Westerkamp, P. Frank, M. Freyberg, M. Salvato, and T. Enßlin. Bayesian Multi-wavelength Imaging of the LMC SN1987A with SRG/eROSITA, 2024a. URL <https://arxiv.org/abs/2410.14599>.
- V. Eberle, M. Guardiani, M. Westerkamp, P. Frank, J. Rüstig, J. Stadler, and T. A. Enßlin. J-UBIK: The JAX-accelerated Universal Bayesian Imaging Kit, 2024b. URL <https://arxiv.org/abs/2409.10381>.
- G. Edenhofer, P. Frank, J. Roth, R. H. Leike, M. Guerdi, L. I. Scheel-Platz, M. Guardiani, V. Eberle, M. Westerkamp, and T. A. Enßlin. Re-Envisioning Numerical Information Field Theory (NIFTy.re): A Library for Gaussian Processes and Variational Inference. *Journal of Open Source Software*, 9(98):6593, 2024. doi: 10.21105/joss.06593.

- A. Ellien, E. Greco, and J. Vink. Evidence for Thermal X-Ray Emission from the Synchrotron-dominated Shocks in Tycho's Supernova Remnant. *The Astrophysical Journal*, 951(2):103, 2023. doi: 10.3847/1538-4357/accc85.
- T. A. Enßlin. Information field theory. *AIP Conf. Proc.*, 1553:184, 2013. doi: 10.1063/1.4819999.
- T. A. Enßlin. Information theory for fields. *Annalen der Physik*, 531(3), 2019. doi: 10.1002/andp.201800127.
- T. A. Enßlin. Information Field Theory and Artificial Intelligence. *Entropy*, 24(3):374, 2022. doi: 10.3390/e24030374.
- T. A. Enßlin, M. Frommert, and F. S. Kitaura. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Physical Review D*, 80:105005, 2009. doi: 10.1103/PhysRevD.80.105005.
- I. Evans, F. Primini, K. Glotfelty, C. Anderson, N. Bonaventura, J. Chen, J. Davis, S. Doe, J. Evans, G. Fabbiano, E. Galle, D. Gibbs, J. Grier, R. Hain, D. Hall, P. Harbo, X. He, J. Houck, M. Karovska, and P. Zografou. The Chandra Source Catalog. *The Astrophysical Journal Supplement Series*, 189:37, 2010. doi: 10.1088/0067-0049/189/1/37.
- F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses: Multimodal nested sampling. *Monthly Notices of the Royal Astronomical Society*, 384(2):449–463, 2008. doi: 10.1111/j.1365-2966.2007.12353.x.
- F. Feroz, M. P. Hobson, and M. Bridges. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, 2009. doi: 10.1111/j.1365-2966.2009.14548.x.
- F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt. Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, 2(1), 2019. doi: 10.21105/astro.1306.2144.
- A. Fowlie, W. Handley, and L. Su. Nested sampling with plateaus. *Monthly Notices of the Royal Astronomical Society*, 503(1):1199–1205, 2021. doi: 10.1093/mnras/stab590.
- P. Frank, R. Leike, and T. A. Enßlin. Geometric Variational Inference. *Entropy*, 23(7):853, 2021. doi: 10.3390/e23070853.
- P. Freeman, V. Kashyap, R. Rosner, and D. Lamb. A Wavelet-Based Algorithm for the Spatial Analysis of Poisson Data. *The Astrophysical Journal Supplement Series*, 138:185, 2008. doi: 10.1086/324017.

- A. Fruscione, J. McDowell, G. E. Allen, N. S. Brickhouse, D. Burke, J. E. Davis, N. Durham, M. Elvis, E. C. Galle, D. E. Harris, D. P. Huenemoerder, J. C. Houck, B. Ishibashi, M. Karovska, F. Nicastro, M. S. Noble, M. A. Nowak, F. A. Primini, A. Siemiginowska, R. K. Smith, and M. W. Wise. CIAO: Chandra's data analysis system. In *SPIE Astronomical Telescopes + Instrumentation*, 2006.
- G. P. Garmire, M. W. Bautz, P. G. Ford, J. A. Nousek, and G. R. Ricker Jr. Advanced CCD imaging spectrometer (ACIS) instrument on the Chandra X-ray Observatory. In *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*, volume 4851, pages 28–44, 2003. doi: 10.1117/12.461599.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis, Third Edition*. Taylor & Francis, 2013. ISBN 9781439840955.
- S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984. doi: 10.1109/TPAMI.1984.4767596.
- F. Guglielmetti, R. Fischer, and V. Dose. Background-source separation in astronomical images with Bayesian probability theory - I. The method. *Monthly Notices of the Royal Astronomical Society*, 396(1):165–190, 2009. doi: 10.1111/j.1365-2966.2009.14739.x.
- M. Habeck. Nested sampling with demons. *AIP Conference Proceedings*, 1641(1):121–129, 2015. doi: 10.1063/1.4905971.
- F. Haberl, U. Geppert, B. Aschenbach, and G. Hasinger. XMM-Newton observations of SN 1987 A. *Astronomy & Astrophysics*, 460(3):811–819, 2006. doi: 10.1051/0004-6361:20066198.
- W. Handley. anesthetic: nested sampling visualisation. *The Journal of Open Source Software*, 4(37):1414, 2019. doi: 10.21105/joss.01414.
- W. Handley. Curvature tension: Evidence for a closed universe. *Physical Review D*, 103(4), 2021. doi: 10.1103/physrevd.103.l041301.
- W. Handley and P. Lemos. Quantifying dimensionality: Bayesian cosmological model complexities. *Physical Review D*, 100(2):023512, 2019. doi: 10.1103/PhysRevD.100.023512.
- W. J. Handley, M. P. Hobson, and A. N. Lasenby. polychord: next-generation nested sampling. *Monthly Notices of the Royal Astronomical Society*, 453(4):4385–4399, 2015. doi: 10.1093/mnras/stv1911.
- R. K. Heilmann, D. P. Huenemoerder, J. A. McCoy, and R. L. McEntaffer. Diffraction Gratings for X-ray Spectroscopy. 2024. URL <https://arxiv.org/abs/2409.02297>.

- E. A. Helder, J. Vink, A. M. Bykov, Y. Ohira, J. C. Raymond, and R. Terrier. Observational Signatures of Particle Acceleration in Supernova Remnants. *Space Science Reviews*, 173 (1-4):369–431, 2012. doi: 10.1007/s11214-012-9919-8.
- E. Higson, W. Handley, M. Hobson, and A. Lasenby. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, 29(5):891–913, 2018. doi: 10.1007/s11222-018-9844-0.
- D. W. Hogg and D. Foreman-Mackey. Data analysis recipes: Using markov chain monte carlo*. *The Astrophysical Journal Supplement Series*, 236(1):11, 2018. ISSN 1538-4365. doi: 10.3847/1538-4365/aab76e.
- S. Hutschenreuter, C. S. Anderson, S. Betti, G. C. Bower, J.-A. Brown, M. Brüggen, E. Carretti, T. Clarke, A. Clegg, A. Costa, S. Croft, C. Van Eck, B. M. Gaensler, F. de Gasperin, M. Haverkorn, G. Heald, C. L. H. Hull, M. Inoue, M. Johnston-Hollitt, J. Kaczmarek, C. Law, Y. K. Ma, D. MacMahon, S. A. Mao, C. Riseley, S. Roy, R. Shanahan, T. Shimwell, J. Stil, C. Sobey, S. P. O’Sullivan, C. Tasse, V. Vacca, T. Vernstrom, P. K. G. Williams, M. Wright, and T. A. Enßlin. The galactic faraday rotation sky 2020. *Astronomy & Astrophysics*, 657:A43, 2022. doi: 10.1051/0004-6361/202140486.
- H. Jeffreys. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London Series A*, 186(1007):453–461, 1946. doi: 10.1098/rspa.1946.0056.
- W. A. Joye and E. Mandel. New Features of SAOImage DS9. In *Astronomical Data Analysis Software and Systems XII*, volume 295, page 489, 2003.
- S. Katsuda, R. Petre, K. S. Long, S. P. Reynolds, P. F. Winkler, K. Mori, and H. Tsunemi. The First X-Ray Proper-Motion Measurements of the Forward Shock in the Northeastern Limb of SN 1006. *The Astrophysical Journal*, 692(2):L105, 2009. doi: 10.1088/0004-637X/692/2/L105.
- S. Katsuda, K. S. Long, R. Petre, S. P. Reynolds, B. J. Williams, and P. F. Winkler. X-Ray Proper Motions and Shock Speeds Along the Northwest Rim of SN 1006. *The Astrophysical Journal*, 763(2):85, 2013. doi: 10.1088/0004-637x/763/2/85.
- A. Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109, 1934. doi: 10.1007/BF01449156.
- D. P. Kingma, T. Salimans, and M. Welling. Variational Dropout and the Local Reparameterization Trick. In *Advances in Neural Information Processing Systems*, volume 28, 2015. doi: 10.48550/arXiv.1506.02557.
- J. Knollmüller and T. A. Enßlin. Metric Gaussian Variational Inference. 2020. doi: 10.48550/arXiv.1901.11033.

- K. Koyama, R. Petre, E. V. Gotthelf, U. Hwang, M. Matsuura, M. Ozaki, and S. S. Holt. Evidence for shock acceleration of high-energy electrons in the supernova remnant SN1006. *Nature*, 378(6554):255–258, 1995. doi: 10.1038/378255a0.
- S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951. doi: 10.1214/aoms/1177729694.
- S. Kumaran, S. Mandal, S. Bhattacharyya, and D. Mishra. Automated classification of Chandra X-ray point sources using machine learning methods. *Monthly Notices of the Royal Astronomical Society*, 520(4):5065–5076, 2023. doi: 10.1093/mnras/stad414.
- J. Li, A. Decourchelle, M. Miceli, J. Vink, and F. Bocchino. XMM–Newton large program on SN1006 – I. Methods and initial results of spatially resolved spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 453(4):3953–3974, 2015. doi: 10.1093/mnras/stv1882.
- T. Liu, A. Merloni, J. Comparat, K. Nandra, J. Sanders, G. Lamer, J. Buchner, T. Dwelly, M. Freyberg, A. Malyali, A. Georgakakis, M. Salvato, H. Brunner, M. Brusa, M. Klein, V. Ghirardini, N. Clerc, F. Pacaud, E. Bulbul, A. Liu, A. Schwobe, J. Robrade, J. Wilms, T. Dauser, M. E. Ramos-Ceja, T. H. Reiprich, T. Boller, and J. Wolf. Establishing the X-ray source detection strategy for eROSITA with simulations. *Astronomy & Astrophysics*, 661:A27, 2022. doi: 10.1051/0004-6361/202141178.
- K. S. Long, S. P. Reynolds, J. C. Raymond, P. F. Winkler, K. K. Dyer, and R. Petre. Chandra CCD Imagery of the Northeast and Northwest Limbs of SN 1006. *The Astrophysical Journal*, 586(2):1162, 2003. doi: 10.1086/367832.
- X. Lu and B. Van Roy. Ensemble sampling. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 2017. doi: 10.48550/arXiv.1705.07347.
- M. Matsuura, M. Boyer, R. G. Arendt, J. Larsson, C. Fransson, A. Rest, A. P. Ravi, S. Park, P. Cigan, T. Temim, E. Dwek, M. J. Barlow, P. Bouchet, G. Clayton, R. Chevalier, J. Danziger, J. De Buizer, I. De Looze, G. De Marchi, O. Fox, C. Gall, R. D. Gehrz, H. L. Gomez, R. Indebetouw, T. Kangas, F. Kirchschrager, R. Kirshner, P. Lundqvist, J. M. Marcaide, I. Martí-Vidal, M. Meixner, D. Milisavljevic, S. Orlando, M. Otsuka, F. Priestley, A. M. S. Richards, F. Schmidt, L. Staveley-Smith, N. Smith, J. Spyromilio, J. Vink, L. Wang, D. Watson, R. Wesson, J. C. Wheeler, C. E. Woodward, G. Zanardo, D. Alp, and D. Burrows. Deep jwst/nircam imaging of supernova 1987a. *Monthly Notices of the Royal Astronomical Society*, 532(4):3625–3642, 2024. doi: 10.1093/mnras/stae1032.
- J. McDowell. Coordinate systems for analysis of on-orbit chandra data paper i: Imaging. *CXC Internal Document, (March 2000)*, 2001.

- J. C. McDowell. X-ray Astronomy and the Analysis of X-ray Data. In *Astronomical Data Analysis Software and Systems XV*, volume 351, page 47, 2006.
- A. Merloni, G. Lamer, T. Liu, M. E. Ramos-Ceja, H. Brunner, E. Bulbul, K. Dennerl, V. Doroshenko, M. J. Freyberg, S. Friedrich, E. Gatzert, A. Georgakakis, F. Haberl, Z. Igo, I. Kreykenbohm, A. Liu, C. Maitra, A. Malyali, M. G. F. Mayer, K. Nandra, P. Predehl, J. Robrade, M. Salvato, J. S. Sanders, I. Stewart, D. Tübing-Arenas, P. Weber, J. Wilms, R. Arcodia, E. Artis, J. Aschersleben, A. Avakyan, C. Aydar, Y. E. Bahar, F. Balzer, W. Becker, K. Berger, T. Boller, W. Bornemann, M. Brüggen, M. Brusa, J. Buchner, V. Burwitz, F. Camilloni, N. Clerc, J. Comparat, D. Coutinho, S. Czesla, S. M. Dannhauer, L. Dauner, T. Dauser, J. Dietl, K. Dolag, T. Dwelly, K. Egg, E. Ehl, S. Freund, P. Friedrich, R. Gaida, C. Garrel, V. Ghirardini, A. Gokus, G. Grünwald, S. Grandis, I. Grotova, D. Gruen, A. Gueguen, S. Hämmerich, N. Hamaus, G. Hasinger, K. Haubner, D. Homan, J. Ider Chitham, W. M. Joseph, A. Joyce, O. König, D. M. Kaltenbrunner, A. Khokhriakova, W. Kink, C. Kirsch, M. Kluge, J. Knies, S. Krippendorf, M. Krumpe, J. Kurpas, P. Li, Z. Liu, N. Locatelli, M. Lorenz, S. Müller, E. Magaudda, C. Mannes, H. McCall, N. Meidinger, M. Michailidis, K. Migkas, D. Muñoz-Giraldo, B. Musiimenta, N. T. Nguyen-Dang, Q. Ni, A. Olechowska, N. Ota, F. Pacaud, T. Pasini, E. Perinati, A. M. Pires, C. Pommranz, G. Ponti, K. Poppenhaeger, G. Pühlhofer, A. Rau, M. Reh, T. H. Reiprich, W. Roster, S. Saeedi, A. Santangelo, M. Sasaki, J. Schmitt, P. C. Schneider, T. Schrabback, N. Schuster, A. Schwoppe, R. Seppi, M. M. Serim, S. Shreeram, E. Sokolova-Lapa, H. Starck, B. Stelzer, J. Stierhof, V. Suleimanov, C. Tenzer, I. Traulsen, J. Trümper, K. Tsuge, T. Urrutia, A. Veronica, S. G. H. Waddell, R. Willer, J. Wolf, M. C. H. Yeung, A. Zainab, F. Zangrandi, X. Zhang, Y. Zhang, and X. Zheng. The SRG/eROSITA all-sky survey: First X-ray catalogues and data release of the western Galactic hemisphere. *Astronomy & Astrophysics*, 682:A34, 2024. doi: 10.1051/0004-6361/202347165.
- T. J. Mitchell and J. J. Beauchamp. Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988. doi: 10.2307/2290129.
- P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006. doi: 10.1111/j.1467-9868.2006.00553.x.
- J. G. Nagy and D. P. O’Leary. Fast iterative image restoration with a spatially varying psf. In *Advanced Signal Processing: Algorithms, Architectures, and Implementations VII*, volume 3162, pages 388–399, 1997. doi: 10.1117/12.279513.
- A. Picquenot, F. Acero, J. Bobin, P. Maggi, J. Ballet, and G. W. Pratt. Novel method for component separation of extended sources in X-ray astronomy. *Astronomy & Astrophysics*, 627:A139, 2019. doi: 10.1051/0004-6361/201834933.

- P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, H. Böhringer, A. Bogomolov, T. Boller, K. Borm, W. Bornemann, H. Bräuninger, M. Brüggen, H. Brunner, M. Brusa, E. Bulbul, M. Buntov, V. Burwitz, W. Burkert, N. Clerc, E. Churazov, D. Coutinho, T. Dauser, K. Dennerl, V. Doroshenko, J. Eder, V. Emberger, T. Eraerds, A. Finoguenov, M. Freyberg, P. Friedrich, S. Friedrich, M. Fürmetz, A. Georgakakis, M. Gilfanov, S. Granato, C. Grossberger, A. Gueguen, P. Gureev, F. Haberl, O. Hälker, G. Hartner, G. Hasinger, H. Huber, L. Ji, A. von Kienlin, W. Kink, F. Korotkov, I. Kreykenbohm, G. Lamer, I. Lomakin, I. Lapshov, T. Liu, C. Maitra, N. Meidinger, B. Menz, A. Merloni, T. Mernik, B. Mican, J. Mohr, S. Müller, K. Nandra, V. Nazarov, F. Pacaud, M. Pavlinsky, E. Perinati, E. Pfeffermann, D. Pietschner, M. E. Ramos-Ceja, A. Rau, J. Reiffers, T. H. Reiprich, J. Robrade, M. Salvato, J. Sanders, A. Santangelo, M. Sasaki, H. Scheuerle, C. Schmid, J. Schmitt, A. Schwobe, A. Shirshakov, M. Steinmetz, I. Stewart, L. Strüder, R. Sunyaev, C. Tenzer, L. Tiedemann, J. Trümper, V. Voron, P. Weber, J. Wilms, and V. Yaroshenko. The eROSITA x-ray telescope on SRG. *Astronomy & Astrophysics*, 647:A1, 2021. doi: 10.1051/0004-6361/202039313.
- R. Preuss, U. Von Toussaint, K. H. Knuth, A. Caticha, J. L. Center, A. Giffin, and C. C. Rodríguez. Comparison of Numerical Methods for Evidence Calculation. In *AIP Conference Proceedings*, volume 954, pages 221–228, 2007. doi: 10.1063/1.2821265.
- D. Pumpe, M. Reinecke, and T. A. Enßlin. Denoising, Deconvolving and Decomposing multi-Dimensional Photon Observations- The D4PO Algorithm. *Astronomy & Astrophysics*, 619, 2018. doi: 10.1051/0004-6361/201832781.
- S. P. Reynolds. Supernova Remnants at High Energy. *Annual Review of Astronomy and Astrophysics*, 46(1):89–126, 2008. doi: 10.1146/annurev.astro.46.060407.145237.
- P. C. Robert. *The Metropolis–Hastings Algorithm*, pages 1–15. 2015. doi: <https://doi.org/10.1002/9781118445112.stat07834>.
- J. Roth, P. Arras, M. Reinecke, R. A. Perley, R. Westermann, and T. A. Enßlin. Bayesian radio interferometric imaging with direction-dependent calibration. *Astronomy & Astrophysics*, 678:A177, 2023. doi: 10.1051/0004-6361/202346851.
- G. B. Rybicki and A. P. Lightman. *Radiative Processes in Astrophysics*. A Wiley-Interscience publication. Wiley, 1991. ISBN 9780471827597.
- R. Salomone, L. F. South, A. M. Johansen, C. Drovandi, and D. P. Kroese. Unbiased and Consistent Nested Sampling via Sequential Monte Carlo, 2023.
- J. S. Sanders, A. C. Fabian, H. R. Russell, S. A. Walker, and K. M. Blundell. Detecting edges in the X-ray surface brightness of galaxy clusters. *Mon. Not. R. Astron. Soc.*, 460(2):1898–1911, 2016. doi: 10.1093/mnras/stw1119.

- M. Sasaki, J. Knies, F. Haberl, C. Maitra, J. Kerp, A. M. Bykov, K. Dennerl, M. D. Filipović, M. Freyberg, B. S. Koribalski, S. Points, and L. Staveley-Smith. First studies of the diffuse x-ray emission in the large magellanic cloud with eROSITA. *Astronomy & Astrophysics*, 661:A37, 2022. doi: 10.1051/0004-6361/202141054.
- N. Schartel and M. Dahlem. Europas Röntgenobservatorium XMM. *Physikalische Blätter*, 56(2):37–42, 2000. doi: 10.1002/phbl.20000560210.
- L. I. Scheel-Platz, J. Knollmüller, P. Arras, P. Frank, M. Reinecke, D. Jüstel, and T. A. Enßlin. Multicomponent imaging of the Fermi gamma-ray sky in the spatio-spectral domain. *Astronomy & Astrophysics*, 680:A2, 2023. doi: 10.1051/0004-6361/202243819.
- D. Schittenhelm and P. Wacker. Nested Sampling And Likelihood Plateaus. 2021. doi: 10.48550/arXiv.2005.08602.
- A. Schöniger, T. Wöhling, L. Samaniego, and W. Nowak. Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence. *Water Resources Research*, 50(12):9484–9513, 2014. doi: 10.1002/2014WR016062.
- M. Selig and T. A. Enßlin. Denoising, deconvolving, and decomposing photon observations. *Astronomy & Astrophysics*, 574:A74, 2015. doi: 10.1051/0004-6361/201323006.
- M. Selig, M. R. Bell, H. Junklewitz, N. Oppermann, M. Reinecke, M. Greiner, C. Pachajoa, and T. A. Enßlin. Nifty – numerical information field theory: A versatile python library for signal inference. *Astronomy & Astrophysics*, 554:A26, 2013. doi: 10.1051/0004-6361/201321236.
- F. D. Seward and P. A. Charles. *Exploring the X-ray Universe*. Cambridge University Press, 2 edition, 2010.
- J. Skilling. Nested Sampling. In *AIP Conference Proceedings*, volume 735, pages 395–405, 2004. doi: 10.1063/1.1835238.
- J. Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4): 833 – 859, 2006. doi: 10.1214/06-BA127.
- J. Skilling. Bayesian computation in big spaces-nested sampling and Galilean Monte Carlo. *AIP Conference Proceedings*, 1443(1):145–156, 2012. doi: 10.1063/1.3703630.
- J. S. Speagle. dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *Monthly Notices of the Royal Astronomical Society*, 493(3): 3132–3158, 2020. doi: 10.1093/mnras/staa278.
- T. Steininger, J. Dixit, P. Frank, M. Greiner, S. Hutschenreuter, J. Knollmüller, R. Leike, N. Porqueres, D. Pumpe, M. Reinecke, M. Šraml, C. Varady, and T. Enßlin. NIFTy 3 - Numerical Information Field Theory: A Python Framework for Multicomponent Signal Inference on HPC Clusters. *Annalen der Physik*, 531(3):1800290, 2019. doi: 10.1002/andp.201800290.

- R. Stephenson. SN 1006: the brightest supernova. *Astronomy & Geophysics*, 51(5):5.27–5.32, 2010. doi: 10.1111/j.1468-4004.2010.51527.x.
- R. Sunyaev, V. Arefiev, V. Babyshkin, A. Bogomolov, K. Borisov, M. Buntov, H. Brunner, R. Burenin, E. Churazov, D. Coutinho, J. Eder, N. Eismont, M. Freyberg, M. Gilfanov, P. Gureyev, G. Hasinger, I. Khabibullin, V. Kolmykov, S. Komovkin, R. Krivonos, I. Lapshov, V. Levin, I. Lomakin, A. Lutovinov, P. Medvedev, A. Merloni, T. Mernik, E. Mikhailov, V. Molodtsov, P. Mzhelsky, S. Müller, K. Nandra, V. Nazarov, M. Pavlinsky, A. Poghodin, P. Predehl, J. Robrade, S. Sazonov, H. Scheuerle, A. Shirshakov, A. Tkachenko, and V. Voron. Srg x-ray orbital observatory - its telescopes and first scientific results. *Astronomy & Astrophysics*, 656:A132, 2021. doi: 10.1051/0004-6361/202141179.
- J. Truemper. The ROSAT mission. *Advances in Space Research*, 2(4):241–249, 1982. doi: 10.1016/0273-1177(82)90070-9.
- J. Vink. Supernova remnants: the X-ray perspective. *The Astronomy and Astrophysics Review*, 20(1), 2011. doi: 10.1007/s00159-011-0049-1.
- J. S. Warren, J. P. Hughes, C. Badenes, P. Ghavamian, C. F. McKee, D. Moffett, P. P. Plucinsky, C. Rakowski, E. Reynoso, and P. Slane. Cosmic-Ray Acceleration at the Forward Shock in Tycho’s Supernova Remnant: Evidence from Chandra X-Ray Observations. *The Astrophysical Journal*, 634(1):376–389, 2005. doi: 10.1086/496941.
- I. Wegener. Simulated Annealing Beats Metropolis in Combinatorial Optimization. *Lecture Notes in Computer Science*, 3580, 2005. doi: 10.1007/11523468_48.
- M. Westerkamp, I. Ovchinnikov, P. Frank, and T. A. Enßlin. Dynamical Field Inference and Supersymmetry. *Entropy*, 23(12):1652, 2021. doi: 10.3390/e23121652.
- M. Westerkamp, I. V. Ovchinnikov, P. Frank, and T. Enßlin. Analysis of Dynamical Field Inference in a Supersymmetric Theory. *Physical Sciences Forum*, 5(1):27, 2022. doi: 10.3390/psf2022005027.
- M. Westerkamp, J. Roth, P. Frank, W. Handley, and T. Enßlin. Inferring Evidence from Nested Sampling Data via Information Field Theory. *Physical Sciences Forum*, 9(1):19, 2023. doi: 10.3390/psf2023009019.
- M. Westerkamp, V. Eberle, M. Guardiani, P. Frank, L. Scheel-Platz, P. Arras, J. Knollmüller, J. Stadler, and T. Enßlin. The first spatio-spectral Bayesian imaging of SN1006 in X-rays. *Astronomy & Astrophysics*, 684:A155, 2024a. doi: 10.1051/0004-6361/202347750.
- M. Westerkamp, J. Roth, P. Frank, W. Handley, and T. Enßlin. Towards a Field-Based Bayesian Evidence Inference from Nested Sampling Data. *Entropy*, 26(11):930, 2024b. doi: 10.3390/e26110930.

- R. Willingale, R. G. West, J. P. Pye, and G. C. Stewart. ROSAT PSPC observations of the remnant of SN 1006. *Monthly Notices of the Royal Astronomical Society*, 278(3): 749–762, 1996. doi: 10.1093/mnras/278.3.749.
- P. F. Winkler, G. Gupta, and K. S. Long. The SN 1006 Remnant: Optical Proper Motions, Deep Imaging, Distance, and Brightness at Maximum. *The Astrophysical Journal*, 585(1):324, 2003. doi: 10.1086/345985.
- P. F. Winkler, B. J. Williams, S. P. Reynolds, R. Petre, K. S. Long, S. Katsuda, and U. Hwang. A high-resolution x-ray and optical study of sn 1006: asymmetric expansion and small-scale structure in a type ia supernova remnant. *The Astrophysical Journal*, 781(2):65, 2014. doi: 10.1088/0004-637X/781/2/65.
- H. Wolter. Spiegelsysteme streifenden Einfalls als abbildende Optiken für Röntgenstrahlen. *Annalen der Physik*, 445(1):94–114, 1952. doi: 10.1002/andp.19524450108.
- G. Zanardo, L. Staveley-Smith, C.-Y. Ng, B. M. Gaensler, T. M. Potter, R. N. Manchester, and A. K. Tzioumis. High-resolution radio observations of SNR 1987A at high frequencies. *The Astrophysical Journal*, 767(2):98, 2013. doi: 10.1088/0004-637X/767/2/98.
- F. Zangrandi, K. Jurk, M. Sasaki, J. Knies, M. D. Filipovic, F. Haberl, P. Kavanagh, C. Maitra, P. Maggi, S. Saeedi, D. Bernreuther, B. Koribalski, S. Points, and L. Staveley-Smith. First Study of the Supernova Remnant Population in the Large Magellanic Cloud with eROSITA. *Astronomy & Astrophysics*, 2024. doi: 10.1051/0004-6361/202348868.