Quantifying sources of uncertainty for convective-scale predictability:

Tracing forecast uncertainty to the origin

Takumi Matsunobu

München 2024

Quantifying sources of uncertainty for convective-scale predictability:

Tracing forecast uncertainty to the origin

Takumi Matsunobu

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

vorgelegt von
Takumi Matsunobu
aus Ibaraki, Japan

München, den 4. Juli 2024

Erstgutachter: Prof. Dr. George C. Craig
Zweitgutachter: Prof. Dr. Markus Rapp
Tag der mündlichen Prüfung: 8. Oktober 2024

Parts of this thesis are included in:

Matsunobu, T., Keil, C. and Barthlott, C. (2022) The impact of microphysical uncertainty conditional on initial and boundary condition uncertainty under varying synoptic control. *Weather and Climate Dynamics*, **3**, 1273–1289, doi: 10.5194/wcd-3-1273-2022.

Matsunobu, T., Puh, M. and Keil, C. (2024) Flow- and scale-dependent spatial predictability of convective precipitation combining different model uncertainty representations. *Quarterly Journal of the Royal Meteorological Society*, **150(761)**, 2364–2381, doi: 10.1002/qj.4713.

Matsunobu, T., Keil, C., and Craig, G. C. (2024) Sampling the impact of model uncertainty and its interaction in a convection-permitting ensemble. *To be submitted.*

Matsunobu, T., Keil, C., and Craig, G. C. (2024) A variance budget to quantify growth and interaction of forecast uncertainties on convective scales. *To be submitted.*

Zusammenfassung

Die numerische Wettervorhersage (NWP) auf konvektiven Skalen hat in den letzten Jahrzehnten große Fortschritte gemacht und ermöglicht bessere Vorhersagen des tagtäglichen Wetters. Die inhärenten Unsicherheiten in NWP-Systemen, die sich aus den Anfangs- und Randbedingungen (IBC), den Modellformulierungen und der Parametrisierung von Prozessen auf der Subgrid-Skala ergeben, stellen jedoch eine Herausforderung für genaue Vorhersagen dar. Die Einführung von Ensemble-Vorhersagesystemen (EPS), die mehrere Vorhersagen mit leicht unterschiedlichen IBCs parallel durchführen, ermöglicht nicht nur probabilistische Vorhersagen, sondern auch eine Schätzung des Vorhersagefehlers durch die Ensemble-Variabilität. Allerdings wird der Fehler oft unterschätzt, weil die Modellunsicherheiten nicht ausreichend dargestellt werden. Die Einbeziehung von unterschiedlichen Darstellungen der Unsicherheit ist für die Verbesserung der Zuverlässigkeit von EPS unerlässlich. In dieser Arbeit wird der Einfluss verschiedener Unsicherheitsquellen auf die Vorhersagequalität und -variabilität in einem operationellen konvektionserlaubenden EPS untersucht. Hierzu werden unterschiedliche konvektive Antriebsbedingungen unterschieden. Der Schwerpunkt liegt auf sommerlicher Konvektion über Deutschland. Die Studie ist in drei miteinander verbundene Teile gegliedert:

Zunächst wird der relative Einfluss einzelner und kombinierter Unsicherheiten, insbesondere derjenige von IBC und mikrophysikalischen Parameterstörungen (MPP), auf den täglichen Niederschlag bewertet. Es wird gezeigt, dass insbesondere die IBC-Unsicherheit die Niederschlagsvariabilität beeinflusst, während die MPP eine sekundäre Rolle spielen. Während schwachem konvektivem Antrieb sind MPP wichtiger als während starkem Antrieb. Die Kombination von MPP mit IBC-Störungen vergrößert die Breite der Vorhersageverteilungen, während der Bereich der Interquartile kaum verändert wird, was auf eine Redundanz der Unsicherheitskombination hindeutet.

Zweitens wird die räumliche Vorhersagbarkeit des stündlichen Niederschlags und die Auswirkung der Unsicherheitsdarstellungen unter Verwendung einer skalenabhängigen Metrik und Beobachtungen untersucht. Es zeigt sich, dass die operationelle konvektiv-skalige EPS des Deutschen Wetterdienstes eine unzureichende räumliche Variabilität aufweist, insbesondere bei schwachen Antriebsbedingungen. Das Hinzufügen von physikalisch basierten stochastischen Störungen (PSP) in der planetaren Grenzschicht reduziert diese räumliche 'Overconfidence'. Die kombinierte Auswirkung von PSP und MPP in Gegenwart von IBC-Störungen auf den räumlichen Fehler und die Streuung scheint additiv zu sein, wobei PSP die Hauptquelle der Modellunsicherheit unter schwachem Antrieb für Konvektion bildet.

Um die Redundanz bzw. Additivität von Unsicherheiten zu quantifizieren, wird drittens eine neuartige „Variabilitätsbudget“-Analysemethode eingeführt. Hier wird die Gesamtvarianz in einzelne Varianzen und deren Korrelationen zerlegt. Die Anwendung dieser Methode auf eine Fallstudie mit variierender Ensemblegröße für IBCs und PSP ergab, dass nur wenige zusätzliche PSP-Member für die Konvergenz der Vorhersage erforderlich sind, falls IBC-Störungen aktiviert sind. Um die Varianzen aus den drei Unsicherheitsquellen nachzuvollziehen, wird die Methode auch auf Ensemblesimulationen angewendet, die einen ganzen Sommermonat umfassen. Es wird gezeigt, dass sowohl PSP als auch MPP in Regionen mit konvektiver Instabilität Variabilität einführen, allerdings auf kleineren Skalen als die, die durch die IBC-Unsicherheit beeinflusst werden. Die Gesamtvarianz wird jedoch aufgrund negativer Korrelationen zwischen den Störungen nicht signifikant verändert. Da die atmosphärische Konvektion von Natur aus hochgradig intermittierend ist und sehr empfindlich auf alle Arten von Modellstörungen reagiert, führen schon geringe Verschiebungen konvektiver Zellen zu gegenphasigen Beziehungen zwischen einzelnen Auswirkungen, was zu einem ineffizienten Anstieg der Varianz führt. Die Korrelationscharakteristiken sind bei wechselndem konvektivem Antrieb ähnlich. Der Einfluß der Modellunsicherheit ist im Vergleich zur IBC-Unsicherheit bei schwachem Antrieb größer. Die stärkere Empfindlichkeit der Modellunsicherheit bei schwachem meteorologischem Antrieb ist eher auf die größere Varianz als auf die Korrelationseigenschaften zurückzuführen. Das „Variabilitätsbudget “läßt sich zur Optimierung eines Ensembledesigns verwenden, um den relativen Einfluß verschiedener Unsicherheitsquellen zu quantifizieren.

Alle in dieser Arbeit vorgestellten Ergebnisse unterstreichen die Notwendigkeit, verschiedener Unsicherheitsquellen in NWP-Systemen zu berücksichtigen. Die neuentwickelte „Variabilitätsbudget“-Analyse ermöglicht eine quantitative Beurteilung der verschiedenen Unsicherheiten. Darüberhinaus zeigt eine strömungsabhängige Betrachtung einen deutlich unterschiedlichen Einfluß von Modellunsicherheiten bei wechselnden atmosphärischen Bedingungen.

Abstract

Numerical weather prediction (NWP) on convective scales has made major advancements in recent decades, enabling more precise forecasts of daily weather. However, inherent uncertainties in NWP systems, stemming from initial and boundary conditions (IBC), model formulations, and parameterisations of subgrid-scale processes, pose challenges for accurate predictions. Introducing ensemble prediction systems (EPSs), which are running multiple forecasts with slightly different IBCs in parallel, enables not only probabilistic forecasting but also an estimate of the forecast error through the ensemble variability. However, the error is often underestimated due to insufficient representation of model uncertainties. Including more uncertainty representations is essential for improving the reliability of EPSs. This thesis explores the influence of various uncertainty sources on convective-scale forecast variability using an operational EPS setting. The growth of variability introduced by model uncertainty representations in the presence of IBC perturbations is examined under varying convective forcing conditions focusing on summer convection over Germany. The study is divided into three interconnected parts:

First, the relative impact of individual and combined uncertainties, specifically IBC and microphysical parameter perturbations (MPP), on daily precipitation is evaluated. It is demonstrated that IBC uncertainty predominantly influences precipitation variability, while MPP play a secondary role with a larger impact during weak convective forcing. Combining MPP with IBC perturbations extends the tail of the forecast distributions while keeping the interquartile range barely changed, suggesting redundancy of the uncertainty combination.

Second, the spatial predictability of hourly precipitation and the impact of uncertainty representations are examined using a scale-dependent metric. It is found that the operational convective-scale EPS at the Deutscher Wetterdienst exhibits insufficient spatial variability, particularly in weak forcing conditions. Adding physically-based stochastic perturbations (PSP) in the planetary boundary layer reduces this spatial overconfidence. The combined impact of PSP and MPP with IBC perturbations on spatial error and spread appears additive, with PSP being the primary source of model uncertainty in weakly forced conditions.

Third, a novel "variability budget" analysis method is introduced to quantify such redundancy or additivity of uncertainties by decomposing total variance into individual variances and their correlations. Applying this method to a case study with varying ensemble size for IBCs and PSP revealed that only few additional PSP members are required

for convergence in the presence of IBC perturbations. This facilitates the optimisation of an ensemble design that includes multiple sources of uncertainty. In order to trace the variances from the three uncertainty sources, the method is also applied to ensemble simulations encompassing an entire summer month. It is demonstrated that both PSP and MPP introduce variability in regions of potential convective instability, but on scales smaller than those represented by IBC uncertainty. However, the total variance is not significantly altered due to negative correlations between the impacts. Since atmospheric convection is highly intermittent in nature and reacts very sensitive to all kinds of model perturbations, only slight displacements of convective cells lead to anti-phase relationships between their individual impacts, leading to an inefficient variance increase. The correlation characteristics are found to be similar in varying convective forcing, but with larger model uncertainty variances relative to that from IBC uncertainty during weak forcing. The stronger sensitivity of weak forcing conditions to model uncertainty is attributable to the larger variance rather than the correlation characteristics.

All results presented in this thesis emphasise the importance of considering various uncertainty sources in NWP systems and strongly suggest their evaluation in a full EPS including IBC uncertainty. The novel variability budget analysis offers a framework for the quantitative assessment of various uncertainties. A flow-dependent evaluation is desired to highlight the variable impact of model uncertainties during varying atmospheric conditions.

Contents

Chapter 1

Introduction

1.1 Numerical Weather Prediction (NWP)

1.1.1 Development of NWP

After humans discovered the equations of fluid motion, the idea of forecasting future atmospheric conditions using them emerged. Bjerknes (1904) settled this idea as an initial-value problem that can be solved with two conditions: the *sufficiently accurate* state of the atmosphere at a given time and the *sufficiently accurate* law according to which the state evolves from a time to another. He successfully enticed people to support expanding observational networks and making a sound basis to realise numerical weather prediction (Kalnay, 2003). However, it remained untried challenge until Richardson (1922) proposed the practical method for the solution, which is numerically integrating the equations of motion. Although his attempt was not successful (Lynch, 2008), he showed the problems that future followers would have to cope with to turn NWP into a reality.

The historical, first success of NWP had to wait until the emergence of the first electronic computers (the Electronic Numerical Integrator and Computer, ENIAC). Charney et al. (1950) computed a reasonable one-day forecast of the two-dimensional vertical mean state of the atmosphere using ENIAC. Although the outcome was quite encouraging for the future of NWP, they had already noticed the potentially great impact of motion at too small scales to be resolved, as well as the impact of divergent flows that are not taken into account in the two-dimensional vorticity equation used for the first forecast. Charney already suggested in his subsequent work shortly coming after that including parameterisations of some subgrid-scale physics to account for those unresolved processes (Charney, 1951). At that time scientists already had a scent that there is an upper limit of predictability due to deficiencies in models' ability to resolve physical processes, which are subjected to practical limitations such as insufficient observations and limited computational resources (Kalnay, 2003). In other words, the limit would be set by failing to fill the two conditions presented by Bjerknes (1904), and if we could continuously improve a model and initial conditions, there would be a boundless improvement. This scent was appeared to be true as forecasts improved, but in theory the dream survived until Lorenz

shows that a limit of predictability is rather imposed by the nature of the atmosphere, even if the model is completely perfect.

Although we have never fully understood the real physics of the atmosphere, solutions to the fluid equations evolve deterministically over time. That deterministic evolution had given hope for an unlimited predictability of the atmosphere. However, Lorenz (1963) showed using simplified thermodynamic equations that the deterministic state is diverging from the trajectories initially nearby following the asymptotic rate. In his following work (Lorenz, 1969a,b), he showed that tiny errors in the atmospheric state (i.e. initial conditions, IC) grow rapidly and eventually spoil an entire forecast. The irremovable uncertainties such as observing difficult variables like divergence and inevitable error due to the mathematical truncation are now considered as a source of forecast failure across scales; and that determines the intrinsic limit of predictability.

The necessity of a probabilistic forecast in order to break down this limit in a deterministic forecast was already foreseen half a century ago. The idea of probabilistic approaches is to tell ourselves to what extent the forecast is uncertain in advance, in other words, how large forecast error is expected. Ideally, estimated uncertainty should match average error. The Monte-Carlo like technique was suggested by Leith (1974), which is currently called "ensemble forecast". The technique involves initialising forecasts with slightly perturbed ICs and covering the space of forecast trajectories that are likely to occur. It also allows for an assessment of the limits of predictability, which largely depend on the atmospheric quantity analysed, the prevailing flow situation, the geographical location considered, as well as the spatial and temporal scale of the predicted phenomena. Following the first success for implementation of an operational ensemble prediction system (EPS) at ECMWF in 1992 (Palmer et al., 1992), EPS has become a global-standard way to perform NWP.

1.1.2 Ensemble technique and sources of uncertainty

EPS is by its nature a technique that has to handle uncertainties in the system. The 'Ensemble' must be generated as a group of forecasts that can represent forecast variability introduced by uncertain factors (Chapter 1 and 18 in Palmer and Hagedorn, 2006). One of those factors is the initial conditions of a model, which can be estimated only with a limited accuracy. Another factor is the boundary conditions that are fed to a model to provide information outside of the model. Moreover, uncertainties are introduced intrinsically by a model itself that can only by differentiating governing equations represent the dynamics and physics of the atmosphere. Those three kinds of uncertainties are referred to as the initial condition uncertainty, boundary condition uncertainty, and model uncertainties, respectively.

Initial condition uncertainty

As the NWP is an initial value problem, the ICs as a source of uncertainty has been of importance from the emerge of predictability study in 1960s (Lorenz, 1963, 1969a; Epstein, 1969). Errors in the initial conditions are attributable to insufficient observations, error

in observations, imperfect assimilation of observational data, and incorrect first guesses used as prior conditions for estimation. Furthermore, it is known that the growth of error depends on the initial condition itself (Slingo and Palmer, 2011). Therefore, appropriate estimation of the initial condition error is crucial for any EPS.

To produce the best estimation of IC for forecasts, data assimilation (DA) techniques are used. DA combines the short-range forecast with observations to generate the IC that has the least error. Within DA algorithms observation errors as well as forecast errors are considered to minimise errors in the generated IC. The estimate of those errors is crucial for generating accurate ICs. While DA was initially used to generate the best ICs for deterministic forecasts, it is now used to generate a plausible distribution of initial states for ensemble forecasts. One method is performing an *Ensemble of Data Assimilation* by repeating a deterministic DA with different observation subsets for each ensemble member (e.g. Bonavita et al., 2016). While this method can take advantage of a pre-existing, deterministic DA system, the computational cost is high and not easily scalable. Another strategy is doing 'ensemble DA' alongside which DA algorithms directly produce a set of ICs by using a statistical approach. In Deutscher Wetterdienst (DWD)'s operational limited-area EPS, for example, the initial conditions are produced using KENDA system (Schraff et al., 2016) based on LETKF (Local Ensemble Transform Kalman Filter). With LETKF, the ensemble distribution made up by ensemble first guesses are transformed using a background forecast error covariance into the new distribution that represents the distribution of the initial condition uncertainty. Since the background forecast error covariance is updated at every DA step in both methods, the produced perturbations represent the up-to-date distribution of uncertainty that reflects flow-dependent uncertainty structure.

Boundary condition uncertainty

Boundary conditions provide information about environment outside a model that is unknown from ICs, such as surface topography and soil moisture, as well as lateral boundaries for a limited-area model. The former type of boundary conditions are surface boundary conditions, providing effects of oceans and lands on the atmosphere. As the surface boundary conditions can continuously constrain the atmospheric state due to their relatively longer time scale of change, accurate surface boundary condition could attain a longer predictability limit (Shukla, 1998). However, it has been often treated as external forcing that does not or hardly change within a forecasting period. One remedy is accompanying a coupled surface model with an atmospheric model such as a land surface scheme in DWD's NWP model (Schulz et al., 2016), sometimes with a perturbed inputs to that coupled model, but not all models have such coupled model because of high development and maintenance cost. Due to the lack of ability to represent uncertainty in the surface boundary condition, we do not include them in consideration in this study.

The latter type of boundary conditions is the lateral boundary condition (LBC). LBCs are provided by a forecast of a parent model, which is often a global model and has a coarser resolution than the limited-area model. Thus LBCs work as essential input, providing large-scale flow patterns encompassing the limited-area model's domain, but suffer from

lower resolution in time and space than the limited-area model. This could matter when the domain of the limited-area model is small and could be more directly constrained by the lateral boundaries (Vukicevic and Errico, 1990). For example, (Tang et al., 2013) shows that the boundary forcing suppresses the spin-up of small-scale convection even if the model's resolution is sufficient to explicitly develop convection on such scales.

Since LBCs are from forecasts of a parent model, they also include forecast error, necessitating perturbations also in LBCs (Vukicevic and Errico, 1990; Laprise, 2003; Hohenegger and Schär, 2007; Vié et al., 2011). One way of producing perturbed LBCs is generating the boundary conditions from different members of global ensemble forecasts. This method can reasonably estimate the temporal increase in uncertainty in LBCs and has been used in a number of limited-area EPS.

Model uncertainty

Uncertainty in model is attributed to unresolved or poorly represented physical processes. One example of the unresolved process is turbulent eddies that occur at scales close to the model's truncation scale. This truncation error often produces too homogeneous states on a grid-point scale and lead to a less active atmosphere due to missing variability. To remedy this problem, model uncertainty representation have been developed to feedback the effect of those uncertainties back to a resolved scale such as SPPT scheme (Buizza et al., 1999; Shutts and Pallarès, 2014), SKEB (Berner et al., 2011a), stochastic convection scheme (Plant and Craig, 2008), and boundary layer perturbation schemes (Kober and Craig, 2016; Hirt et al., 2019; Clark et al., 2021).

Model uncertainty can also be attributed to poor representation of physics. An easy-to-understand example is cloud and precipitation particles in bulk cloud microphysics schemes. Those microphysics schemes classify water particles into handful of categories such as cloud water, rain water and cloud ice, and predict quantities proportional to their moments such as size distributions. The predicted quantities are converted to other categories based on pre-defined conversion rates. While the state of water in the atmosphere is a spectrum and unable to be so simply categorised, many operational forecasting systems have used those schemes not mainly due to the limitation of computational resources, but also due to the fact that we still don't know how those microphysical processes act and can be modelled. Furthermore, parameters in microphysical schemes are often fixed in time and space. The incomplete grasp of the physical processes in the real world renders the inherent flaws equally applicable to parameterisations of other processes. There may be processes still not discovered or modelled. Since rationale to model a representation of this type of model uncertainty is not well established, the impact of the model uncertainty has been investigated by performing sensitivity experiments by making ensembles with multiple models or physics schemes (e.g. Cintineo et al., 2014; Marinescu et al., 2021) and parameter perturbation experiments with a single physics parameterisation scheme (e.g. Seifert et al., 2012; Igel and van den Heever, 2017b; White et al., 2017). In many of studies the first objective was to identify important factors that we should invest more resources to improve the accuracy of the factor by carrying out observation campaigns and assign

additional computing resources for better resolution of that scheme. Recently stochastic representation of this type of model uncertainty by perturbing parameters has emerged, for example the SPP scheme (Ollinaho et al., 2017; Lang et al., 2021) and MPP (Thompson et al., 2021). The advantages of the stochastic approach are that spatiotemporal heterogeneity can be introduced and the overall impact can be quantified without significantly increasing the ensemble size. However, the detailed investigation using those stochastic parameter perturbation scheme is still underway.

1.2 Predictability of Convection

The previous section provided a general overview of the development of NWP and the treatment of uncertainty. There are a lot of evidence that different weather events dominate weather variability at various temporal and spatial scales (as in Fig. 1.1). Furthermore, these events set the limit of practical predictability for their respective scales. To better understand the challenges in weather forecasting, it is necessary to discuss specific events and their corresponding scales.

This thesis specifically explores predictability of short-range convection forecasts over land, which are of particular interest due to the potentially large economic impact of severe convective precipitation in Germany. Convective events, such as thunderstorms and heavy rainfall, can cause substantial damage to infrastructure, agriculture, and human life. Reliable forecasting of these events is crucial for mitigating their negative effects. As a basis for the study, this section outlines studies about predictability of convection.

1.2.1 Synoptic- and convective-scale predictability

In short- and medium-range weather forecasting, the most important distinction in the practical predictability limit is the scale separation between synoptic and convective scales, which are dominated by different physical processes. Synoptic scales, which refer to large-scale weather systems spanning hundreds to thousands of kilometres (Fig. 1.1), are dominated by large-scale waves such as Rossby waves and cyclonic systems. These systems can be approximated as mostly in the hydrostatic balance, meaning that the typical length of their vertical motion is significantly small compared to their horizontal motion. The resolution of current global NWP models, typically a couple of tens kilometres, is high enough to explicitly resolve these synoptic-scale phenomena. Additionally, the density of observations in the mid-latitude of the northern hemisphere is sufficient to adequately represent these events. As a result, the forecast skill limit for synoptic-scale weather systems can extend beyond one week (Buizza and Leutbecher, 2015).

At convective scales, on the order of 10 kilometres (Fig 1.1), convective systems such as thunderstorms are dominant particularly during summer. In deep convection, uplift wind spans from the surface layer up to 10 km heights, making the scale of vertical momentum transport comparable to the scale of horizontal motion. As a result, equations of three-dimensional fluid motion need to be computed for convective-scale forecasting.

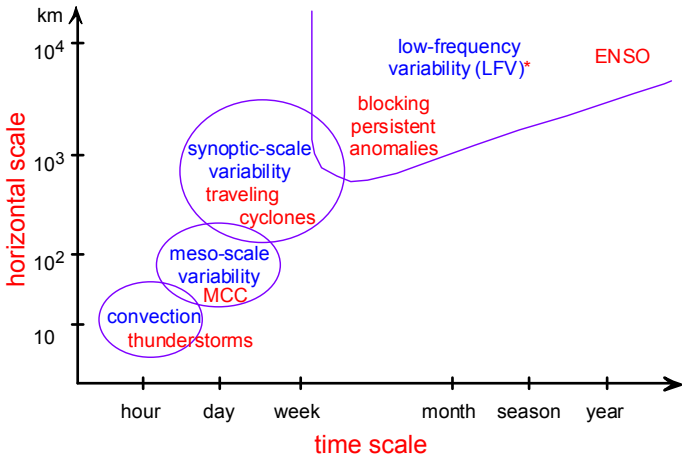

Figure 1.1: Characteristic time and spatial scales of observed atmospheric phenomena. Adopted from Craig et al. (2005).

This three-dimensional motion can no longer be resolved by global NWP models with grid spacing of tens of kilometres. Therefore, models with grid spacing of a few kilometres are required to accurately capture convective-scale phenomena. The rapid growth in computing capability has allowed for the development of models with higher resolution and more intricate parameterisations. These models, called convection-permitting models, have been developed as limited-area models encompassing a target region of forecasts to save computational resources.

Convection-permitting models have led to a step change in rainfall forecasting (Clark et al., 2016). Substantial progress has been achieved with improved data assimilation techniques introducing new types of observations and a better representation of atmospheric convection dynamics through finer grid spacing that made it viable to turn off the parameterisation of deep convection. As a result, convection-permitting models provide on average more spatially precise predictions compared to coarser-resolution models. However, despite these significant advancements, forecasting skill for convective scales is still low compared to that for synoptic scales (Hohenegger and Schär, 2007). One reason is strong non-linearity and local, intermittent character of convection, which leads a slight displacement of convection in time and space to a complete miss forecast. Second, observation networks are too sparse to resolve convective events and we don't have enough observations to constrain model parameters related to convection. Furthermore, convective activity and its predictability is strongly modulated by large-scale flow conditions encompassing the region. Thus, it is necessary to further address sources of uncertainties on forecasting convection by considering their flow and scale dependence. Assessment of predictability using perturbations in IC and LBC has been performed (e.g. Kühnlein et al., 2014) and those have been operationally implemented. This study focuses on unexplored parts, model uncertainties, by combining IC and LBC uncertainties into the form of initial and boundary condition (IBC) uncertainty and investigating the relationship between this

and model uncertainties.

1.2.2 Sources of model uncertainties for convective-scale forecasts

To incorporate model uncertainties in EPS using convection-permitting models, approaches analogous to those in global forecasts have been employed. Unresolved, subgrid-scale physical processes are being considered by applying stochastic perturbation methods, whereas structural uncertainties arising from the formulation of physical processes are frequently represented through multiple-physics or perturbed parameter approaches (e.g. Berner et al., 2017; Fleury et al., 2022; Roberts et al., 2023, and references therein). Despite these advancements, current EPSs with convection-permitting models are often under-dispersive for near-surface variables (e.g. Bouttier et al., 2012; Raynaud and Bouttier, 2017), highlighting the ongoing challenge of adequately representing uncertainty at these scales. In this study, we consider two sources of model uncertainty that have been considered primarily important sources (e.g. Cintineo et al., 2014; Johnson and Wang, 2020; Hermoso et al., 2021): turbulence in the planetary boundary layer and parameter uncertainty in a microphysical scheme.

Planetary boundary layer turbulence as well as cloud microphysics and their interaction with aerosols are key physical processes that represent important sources of model uncertainty in forecasts of convection (Clark et al., 2016). The initiation of convection is predominantly linked to boundary layer processes, but those processes are not fully resolved due to their intrinsic small scales. In many current boundary layer schemes the turbulent processes are represented by a mean state within a grid box. This leads to insufficient small-scale variability and inhibits or delays the initiation of convection especially when convective forcing to trigger convection is missing (Kühnlein et al., 2014). In this study we apply the physically-based stochastic perturbation (PSP) scheme (Kober and Craig, 2016; Hirt et al., 2019) that adds perturbations to the tendencies of T, w, and q in the boundary layer to increase turbulence. Independently Clark et al. (2021) proposed a similar stochastic boundary layer scheme representing turbulent eddies as random events following a Poisson distribution inside a bulk model of the convective boundary layer. The variability in precipitation rates introduced by the stochastic boundary layer scheme is beneficial for better detecting potential risk of flooding, whereas the variability in spatial distribution "fills the gap" left by insufficient representation of uncertainty by perturbations in initial and lateral boundary conditions (Flack et al., 2021).

Microphysical processes constitute another major source of model uncertainty. Among the large number of poorly constrained parameters used in current bulk microphysics schemes, cloud condensation nuclei (CCN) concentrations (Barthlott and Hoose, 2018; Glassmeier and Lohmann, 2018; Dagan and Stier, 2020) and the shape of the cloud droplet size distribution (CDSD, Igel and van den Heever, 2017a,b) are known to be potentially influential parameters in precipitation forecasts. Thompson et al. (2021) and Barthlott et al. (2022b,a) show that both CCN and CDSD uncertainty have non-negligible impact on accu-

mulated precipitation using stochastic and perturbed parameter approaches, respectively. They also find that the combination of CCN and CDSD uncertainties has a larger impact than their univariate impact, although the synergistic impact is smaller than the simple sum of them. Such microphysical parameter perturbations (MPP) are also incorporated to assess the sensitivity of convective-scale forecasts to moist processes.

1.2.3 Flow- and scale-dependence of predictability of convection

Even if forecast uncertainty is sensitive to sources of those uncertainties, it does not matter if the response to them is constant or linear. Unfortunately, many studies have proven that the response behaves quite non-linearly, heterogeneously, and flow-dependently (e.g. Lorenz, 1963; Slingo and Palmer, 2011; Buizza and Leutbecher, 2015). The rate of convection is controlled by a variety of factors. First is the contribution to generation of instability by cooling of the troposphere, mainly by dynamically-driven ascent, but with some contribution from radiation. The cooling aloft is complemented by heating and moistening of the boundary layer by surface fluxes or advection, leading to the creation of convective available potential energy (CAPE). However, the presence of CAPE does not guarantee the occurrence of convection. Triggering by mesoscale and local features is often required to overcome convective inhibition associated with a capping inversion at the top of the planetary boundary layer. Triggering features include convergence lines and other boundary layer structures, but can also include perturbations from previous generations of convective clouds such as outflow boundaries and gravity waves. It is a challenge to identify characteristics of the meteorological situation that impact predictability.

The impact of the dynamical forcing of convection in mid-latitudes is often described in terms of strong or weak forcing. Typically this is identified based on the presence or absence of synoptic or mesoscale dynamical features that can drive ascent and the creation of CAPE. However, the identification of such features is usually subjective and it is useful to have a more precisely defined measure of the influence of the convective environment. For this purpose Done et al. (2006) proposed the convective adjustment timescale τ_c, which is a measure of the extent to which the convection is in equilibrium with the larger-scale forcing. The timescale is defined as the ratio of CAPE to the rate of its removal:

$$\tau_c = \frac{\mathrm{CAPE}}{\frac{d\mathrm{CAPE}}{dt}}. \tag{1.1}$$

If the inhibition of convection is weak and triggering disturbances are plentiful, convection will occur whenever instability is present and CAPE will be consumed as fast as it is created, keeping τ_c close to 0. The amount of convection, measured by say mass flux or precipitation, is in equilibrium with the forcing processes creating instability. On the other hand, if the inhibition of convection is strong and the triggering disturbances are weak or absent, CAPE can accumulate and potentially reach large values, i.e. convection is in non-equilibrium and τ_c often reaches larger values. Equilibrium conditions often coincide with strong forcing because dynamical ascent can weaken inversions while widespread convection provides an

abundance of triggering disturbances to initiate new storms. But it is also possible that a very strong capping inversion or other source of convective inhibition will prevent the convection from reaching equilibrium and the strong forcing will lead to rapid increases in CAPE. The convective adjustment time scale of Done et al. (2006) uses an estimate of the ratio of CAPE to its rate of reduction by convection as a quantitative measure of the degree of equilibrium, as well as of synoptic forcing required for such equilibrium. This measure has been used by a number of studies to assess the flow-dependence of different aspects of convection (Done et al., 2012; Craig et al., 2012; Flack et al., 2018; Keil et al., 2019; Bachmann et al., 2020). For instance, Keil et al. (2014) and Kühnlein et al. (2014) show that the summertime precipitation forecast skill during the non-equilibrium regime in central Europe is below the average for all days of the convective season. The larger forecast uncertainty under non-equilibrium situations due to microphysical uncertainty is reported for Central Europe (Barthlott et al., 2022b,a) and for Eastern CONUS (Chen et al., 2023).

The predictability of weather forecasts also depends on the considered spatial and temporal scale. Generally, the forecast skill limit is considerably longer for spatially and temporally averaged fields (Buizza and Leutbecher, 2015). Over the past decade, spatial metrics besides simple averaging have become widely used to measure forecast skill of convection-permitting models acting on kilometre scales and to estimate its uncertainty (Ebert, 2009; Frogner et al., 2019; Casati et al., 2022). The scale at which convection-permitting models demonstrate reasonable skill remains large compared to their horizontal grid spacing and the typical extent of convective rain. Reasonable forecast skill for 24-h forecasts is achieved at scales of several tens to over a hundred kilometres for models (Mittermaier et al., 2013; Schwartz et al., 2009; Keil et al., 2020). A widely applied spatial measure to inspect the scale dependence is the Fractions Skill Score (FSS, Roberts and Lean, 2008) that relaxes the pointwise comparison, introduces a spatial tolerance and rewards the location proximity of forecast convective cells and observed ones in increasingly larger surroundings.

Some studies quantifying individual impact of boundary-layer and microphysical uncertainties suggest potential benefit of combining representations for these two uncertainty sources. Using a multiple-physics approach, Cintineo et al. (2014) find that the contributions of perturbing boundary-layer and microphysics schemes act on various scales. Boundary layer schemes have a greater spatial impact at the early stage of convection, followed by an increasing impact of microphysical uncertainty. Keil et al. (2019) inspect the relative impact caused by stochastic perturbations in the planetary boundary layer, varied assumptions of CCN concentration and soil moisture heterogeneity on the uncertainty evolution of hourly precipitation rates. While the total amount of daily precipitation is hardly changed by the different perturbation approaches (less than 5%), the spatial variability of precipitation exhibits clear differences. The stochastic boundary-layer perturbations lead to the largest spatial variability from initial time onwards with an amplitude comparable to the operational ensemble spread. Similarly, perturbed aerosol concentrations impact spatial precipitation variability shortly after model initialisation, but to a smaller degree.

The presented importance of flow and scale dependence of predictability raised two

research questions in this thesis about model uncertainty impact on convective-scale forecasts. The first one is understanding **How does the relative impact of individual and combined uncertainties vary under different forcing regimes?** The second question is related to the scale dependence, asking **At what spatial scales do uncertainties impact?**

1.3 Methods to quantify uncertainty

Despite the positive impact of individual representations on improving underdispersion as presented above, it is often found that combining multiple uncertainty representations results in smaller variability than a simple sum of single source variability. The complex interplay between uncertainty impacts slows down or suppresses the growth of forecast variability and leads to a let-down spread increase given the length of forecast period. This has been explained as 'filling the gaps' (Flack et al., 2021), 'compensating effect' (Baur et al., 2022), or simply 'generating not enough/sufficient spread' (Jankov et al., 2019; Hermoso et al., 2021; Frogner et al., 2022), providing qualitative insights for further developments.

Coincidently, the working group of ECMWF Workshop on Model Uncertainty in May 2022 reported the need for further assessments of combinations of model uncertainty schemes, with the aim of identifying a set of the most effective model uncertainty representations for high resolution EPSs (i.e. R6 and R7 in Plant et al., 2022). There is a need for a method that defines a perturbation design to quickly and efficiently improve convective-scale forecast variability. Thus, having a framework to quantify how the impact evolves, how interacts, and why not gain variance as effectively as expected is useful. A statistical parameter sensitivity analysis appears effective to tackle these questions. This section introduces methods commonly used for investigating parameter sensitivity.

By its nature, NWP is a model solving discretised physical equations of the atmosphere. These equations are responding to inputs such as initial conditions and parameter values in physics parameterisation schemes and producing the model output. As discussed former in this section, this model cannot avoid being subject to three kinds of uncertainties originated in those inputs, as well as responses to single uncertainty can change with different conditions due to other uncertainties. A classical method to sample uncertainty impact that perturbs one parameter at one time around a fixed control configuration is insufficient in this case. Parameter sensitivity analysis intended for exploring high-dimensional parameter input space is making it possible to objectively prioritise sources of uncertainties in order of importance for outputs.

Here I would like to introduce methods that intend to screen the effect of large number of input parameters and to quantify the relative importance of them to the model output. Presumably most common methods in the last 10 years is the Sobol's method (Sobol', 1990). The method involves the Monte-Carlo sampling to estimate the PDF of a target variable when one or group of parameters are fixed at a certain point in the parameter space. This allows us to get a full view of the distributions of the target variable and to

understand the various statistics of the sensitivity conditional on any combinations of fixed or perturbed parameters. One proposed analysis of the statistical property of the sensitivity is to decompose the total ensemble variance into the sum of variance contributions due to individual perturbed parameters, called the first-order Sobol's index S (Sobol', 1990; Homma and Saltelli, 1996), which shows how much of the total variance can be reduced by holding a parameter constant, as defined:

$$S_i = \frac{Var(E[f(X)|X_i])}{Var(f(X))},$$

where $f(X)$ is the model output and X_i is the ith input parameter. $Var(E[f(X)|X_i])$ quantifies the variance of the ensemble mean of $f(X)$ given by perturbing the parameter X_i, thus S_i indicates how much the ensemble mean is stabilised if the parameter X_i is accurately known and thus fixed. While this index is very useful to understand which parameter makes the ensemble mean most unstable, it does not provide detailed information of the interaction between parameters. Additionally, the indices are designed for identifying the most influential parameter on reducing the variance of the mean, rather than on a contribution to increasing the ensemble variability, which is of interest to mitigate the underdispersion issue in EPS using convection-permitting models. Furthermore, it requires huge sample size of optimally more than 10,000 for each perturbed parameter (Iooss and Lemaître, 2015), which is almost impossible for NWP models. To complement that difficulty in sampling with a decent sample size, statistical surrogate models are usually used to quantify the impact of uncertainty e.g. the Polynomial Chaos expansion (Wimmer et al., 2022) and Gaussian process emulator (Ji et al., 2018; Ryan et al., 2018; Wellmann et al., 2020). Despite that surrogate approach is an effective way to generate thousands of simulations with a feasible computing effort, it deeply relies on a training dataset and method to build the surrogate. It is might be able to well reproduce the best-fit behaviour for the training dataset on average, but could have trouble with sharping the flow dependence. It is desired to devise a sampling strategy that can be feasible for direct numerical simulations.

Another widely used method is Morris method Morris (1991) that proposes an efficient screening strategy to identify parameters that have trivial impact, i.e. that can be fixed without changing the model outcome (e.g. Covey et al., 2013; Morales et al., 2019; Posselt et al., 2019; Wimmer et al., 2022). The method screens the parameter space by sequentially perturbing inputs "One at a Time", but with changing a starting point of a input perturbation. This involves perturbing one parameter after another and takes the change in the model output as the impact given by the perturbed parameter. The advantage of this method is significantly cheaper computational cost than the Sobol's method, but with the disadvantage in less precision of the parameter's impact. Therefore, the sensitivity indices for the Morris method were designed to estimate relatively simple statistical properties such as the mean and standard deviation of the impact. So far the method has been used to rank the importance of the parameters for roughly identifying parameters that can reduce total variability (Campolongo et al., 2007; Covey et al., 2013; Ji et al., 2018; Morales et al., 2019).

Those methods were developed used to retrieve parameters influential on the variability of the ensemble mean, in order to 'reduce' uncertainty in a model i.e. to identify a parameter to be fixed. However, there is a room for further application of them in identifying parameters that 'increase' variability for improving the underdispersion problem. If a 'budget' of variability due to sources of uncertainties is visualised, it can tell us which combination of uncertainty representations is redundant or efficient. Whereas the budget analysis of the departure of the ensemble mean from observations was developed by Rodwell et al. (2016), it focuses more on separating effects embedded in an EPS as a whole i.e. biases, forecast and observation uncertainties. As far as the author knows, no analysis framework considering the interactions between uncertainty impact has been developed. As described above, the Sobol's and Morris methods seem promising as this framework, but both have own deficiencies. Developing a new framework to quantify a budget of variability inspired by these method will help investigate the uncertainty growth and interactions.

Quantifying the sensitivity in a system that has a large degree of freedom like a NWP system suffers from inevitable sampling uncertainty. Convergence behaviour of forecast distributions have been investigated by bootstrapping a large ensemble perturbed by IBCs (Craig et al., 2022). This study presents that ensemble standard deviations asymptotically converge with sampling error scaling proportional to the inverse square-root of ensemble size and necessary ensemble size is different among variables that have varying distribution shapes. On the other hand, the behaviour for model uncertainty has apparently not been explored yet. Model uncertainty representations are thought to be active within bespoke weather processes, suggesting a possibility of distinct convergence behaviour from that of IBC. Furthermore, sampling uncertainty of interactions in the context of NWP is also waiting for a first investigation.

These facts posed the third research question building a basis for further exploration: **Does model uncertainty and its interactions converge as fast as IBC?** Addressing this question let us take a step forward to an original interest of quantifying the variability growth, which answers the fourth question: **How does forecast uncertainty from different sources evolve and interact over time?**

1.4 Research questions and outline

A lot of efforts have been taken to develop convective-scale EPS as we use today, as well as to understand how a source of uncertainty plays a role in varying flow conditions and at different scales. However, understanding how multiple uncertainties interplay in varying flow situations, spatial scales, and times of day remains elusive. In particular, the impact of these uncertainties on convective-scale forecasts warrants further investigation. This motivates this thesis, which aims to contribute to knowledge for improving forecast reliability by addressing the overarching question of **How do different sources of uncertainty influence the variability of convective-scale forecasts in various flow situations?**

Addressing this question involves answering the four questions posed, which are ar-

ranged to trace uncertainty from large-scale averaged impact towards small scales close to where the impact originates:

1. **Flow dependence**: How does the relative impact of individual and combined uncertainties vary under different convective forcing regimes?

2. **Scale dependence**: At what spatial scales do uncertainties impact?

3. **Model uncertainty convergence** Does model uncertainty and its interactions converge as fast as IBC?

4. **Uncertainty evolution and interaction**: How does forecast uncertainty from different sources evolve and interact over time?

This thesis focuses on investigating summer convection over Germany. The region clearly shows domination of distinct convective flow regimes, and DWD can provide a state-of-the-art convection-permitting model and realistic representations of IC and LBC uncertainties for the region, which are essential for this study. Model uncertainty representations for the boundary layer turbulence and microphysical parameters were available in the model.

Three parts comprise this dissertation. The first part (Chapter 3) evaluates the daily precipitation impact of MPP, in presence of IBC uncertainty. This study complements the assessment of Keil et al. (2019); Barthlott et al. (2022b) to evaluate the combined impact of MPP with IBC variability on a selection of August 2020 case studies. Particular emphasis lies on changes in variability due to presence of other perturbations.

The second part (Chapter 4) investigates the impact on finer temporal and spatial scales. Long-term simulations from the Deutscher Wetterdienst are used to assess the combined impact of IBC perturbations and PSP. The FSS technique is employed to study how adding the PSP scheme gains the spatial variability of hourly precipitation. Additionally, representative case studies are used to showcase the systematic impact of adding PSP and further demonstrate the combined MPP impact with IBC and PSP perturbations.

The third part consists of two chapters and introduces a novel framework to investigate the sensitivity of variance to perturbations. Chapter 5 explores a variance increase and convergence given by IBC and PSP. Sampling uncertainty of variances and correlations is quantified as a function of the number of IBCs and random fields used for the PSP scheme, employing a bootstrapped resampling approach (Craig et al., 2022; Tempest et al., 2023). These analyses compare for the first time the variability growth of IBC and model uncertainty variances and convergence behaviour of their interactions. By utilising a suggested optimal ensemble design from this work, following Chapter 6 then evaluates a one-month ensemble dataset for variance and correlations of IBC, PSP and MPP. The proposed method allows us to trace the impact and interactions of the uncertainties to their origin.

The outline of this dissertation is as follows: Chapter 2 gives an introduction of the NWP model used in this study along with the various uncertainty representation methods.

It further defines the flow-dependent and scale-dependent metrics employed. The last step lays the theoretical foundation for a newly developed sensitivity analysis method. Chapter 3 is the first part of the result analysing the impact of uncertainties on daily accumulated precipitation. Chapter 4 focuses on changes in spatial distributions of hourly precipitation. Chapter 5 constitutes the former part of the third part and examines the convergence behaviour of model uncertainty and its interactions. Chapter 6 traces the evolution of uncertainties to the origin. Finally, Chapter 7 presents the main conclusion and the future research directions.

Chapter 2

Model and Methods

Chapter 2 introduces simulation techniques and analysis methodologies employed to address this study's research questions. Section 2.1–2.2 provides an overview of the model and perturbation approaches used for performing ensemble simulations to address the research questions. Subsequent sections detail three key analysis measures: Section 2.3 introduces the method for identifying convective flow regimes, used throughout the study to identify flow dependence of the impact. Section 2.4 presents and explain Fractions skill scores (FSS; Roberts and Lean, 2008), a method for investigating the scale-dependence of impacts. Section 2.5 offers a theoretical foundation for the newly proposed variance-based method designed for investigating budget of variance given by sources of uncertainties.

2.1 ICON-D2

All numerical simulations are performed with the ICON (ICOsahedral Non-hydrostatic) model in its limited-area mode ICON-D2, which is used in operational weather forecasting at Deutscher Wetterdienst (DWD) since February 2021 (D. Reinert et al., 2021). Given our aim to assess the practical predictability of convective weather using a state-of-the-art EPS, we needed to keep up with the evolution of the ICON model and its associated ICON-D2-EPS implemented by DWD. This necessitated updating the model and dataset used in the study. Thus, only a general description of the ICON model will be provided in this section. Detailed configurations, tailored to specific datasets, will be introduced in each subsequent results chapter.

ICON employs an unstructured icosahedral-triangular Arakawa-C grid in the horizontal direction, formed by spherical triangular cells that seamlessly cover a simulation domain. As described in Zängl et al. (2015), its dynamical core is based on the non-hydrostatic equations for fully compressible fluids. The prognostic variables are the edge horizontal wind speed, vertical wind speed, air density, virtual potential temperature, mixing ratios and, when using the two-moment microphysics scheme (Seifert and Beheng, 2006), the number density of hydrometeors. Time integration is performed using a two-time level predictor-corrector scheme. The ICON-D2 domain covers Central Europe with a grid

spacing of 2 km (542,040 grid cells roughly encompassing 1400 km x 1600 km) and 65 vertically discretised layers from the ground up to 22 km above mean sea level. Figure2.1 shows the simulation domain of ICON-D2 with the model orography. Its prognositc domain covers the entire area of Germany, Austria, Switzerland, Denmark, Netherlands, Belgium, Luxembourg, Slovenia and parts of the neighbouring countries. In this study, Hourly ICON-D2 output data is interpolated onto an uniform, rotated pole coordinate consisting of 651 x 716 grid points (466,116 in total) with a grid spacing of 2.2 km. The evaluation domain in defined based on Baur et al. (2018).

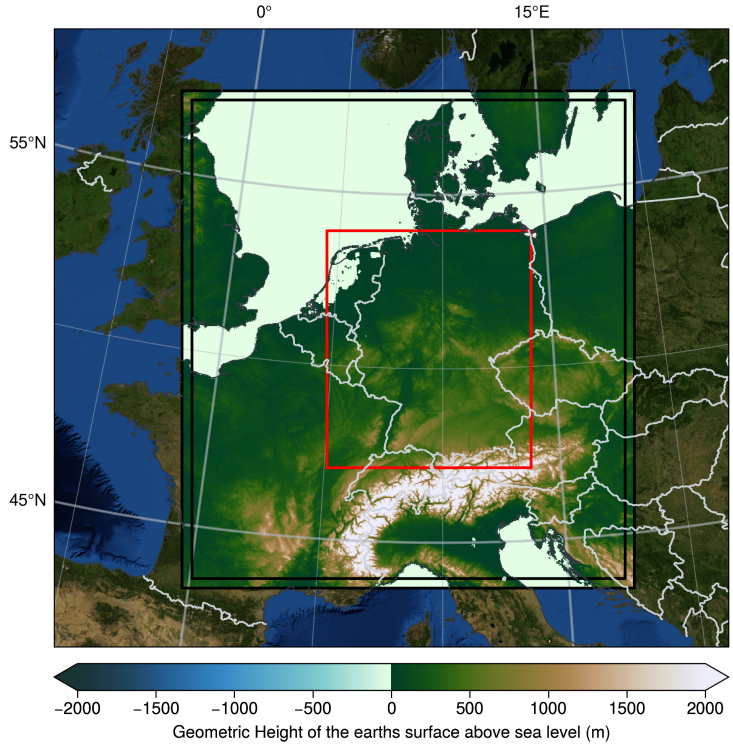

Figure 2.1: ICON-D2 domain (outer black rectangle) and model orography. The inner black rectangle indicates the fully prognostic domain where the lateral boundary conditions are not interpolated nor nudged. The red rectangle shows the German domain used throughout evaluations.

2.2 Uncertainty representation

Representing uncertainty is a key basis for answering research questions in this thesis. This section provides the implementation overview of uncertainty representations in the ICON-D2 model.

2.2.1 Initial and boundary condition perturbations

IC and LBC are the backbone in limited area modelling and represent a major source of forecast uncertainty. In the operational ICON-D2 ensemble prediction system (ICON-D2-EPS) at DWD, IC uncertainty is provided by the ensemble data assimilation system ICON-D2-KENDA (Kilometer-scale ENsemble Data Assimilation Schraff et al., 2016). In the 40-member ICON-D2-KENDA the model state is updated hourly by assimilating observations into 1-hour first guess forecasts. In 2021 only conventional observations (synoptic stations, radiosondes, wind profilers and aircrafts) were operationally assimilated.

The uncertainty representation in LBCs stems from ensemble forecasts generated by the coarser grid model: ICON-EU, nested within the global ICON model. The global ICON-EPS has a horizontal grid spacing of 40 km (26.5 km since November 2022). An ICON-EU nest simulation is embedded online in the global ICON simulation and covers the entire Euro-Atlantic region with half the grid spacing. The ICON-EU ensemble provides the ICON-D2 LBCs. Forecast variability in the global ICON-EPS and ICON-EU-EPS is attained by 40-member IC perturbations generated by the ensemble data assimilation with an assimilation cycle of 3 hours, and by ensemble physics perturbations where a random combination of tuning parameters is set for each of the ensemble members and fixed throughout the forecast horizon. As in DWD's operational setup (D. Reinert et al., 2021), ICON-EU ensemble forecasts initialised three hours before the initialisation time of the ICON-D2 ensemble are used. Therefore, the LBCs are updated hourly using the ICON-EU-EPS output at lead times 3—27 hours. Since we primarily focus on the impact of model uncertainties we consider the impact of IC and LBC uncertainty together and call it IBC uncertainty.

In this study, ensemble simulations were performed closely following the configurations of operational ICON-D2-EPS. Operational analyses from ICON-D2-EPS archived in DWD provided ICs. Note that ICON-D2-EPS was in a preoperational phase until February 2021, and its data were used for the Case Study v1 (Sect. 3.2.1). LBCs were obtained from operational ICON-EU-EPS data. Forecast experiments employed the first 20 members of 40 ICs and 40 LBCs. Throughout the study, the number of IBCs used to make IBC perturbations remained fixed at 20.

2.2.2 Physically based stochastic perturbations

Sub-grid scale uncertainty in the planetary boundary layer is represented here using the PSP scheme presented by Kober and Craig (2016) and revised by Hirt et al. (2019). The rationale of the PSP scheme is to improve the coupling between sub-grid variability and convective initiation in km-scale models. In convection-permitting models, such as the ICON-D2 model, deep convection is explicitly represented. However, processes that lead to convective triggering often occur below the grid scale and are not sufficiently parameterised. These sub-grid processes include boundary layer turbulence, sub-grid orography and density currents resulting from convective downdrafts. The PSP scheme is re-introducing the missing effects of boundary layer turbulence, most influential among the three effects,

in a physically consistent manner. The revised PSP scheme contains modifications for added tendencies: an autoregressive, continuously evolving random field, a limitation of the perturbations to the boundary layer that removes artificial convection at night, and a mask that turns off perturbations in precipitating columns to retain coherent structures (Hirt et al., 2019).

The perturbations are given by adding stochastic disturbances on the model tendencies of temperature, specific humidity and vertical wind within the planetary boundary layer following:

$$\partial_t \Phi|_{PSP} = \alpha_{tuning} \eta \frac{1}{\tau_{eddy}} \frac{l_{eddy}}{5\Delta x} \sqrt{\overline{\Phi'^2}} \tag{2.1}$$

where $\Phi \in T, qv, w$. η is a spatio-temporal structure of random eddy field correlated with a time scale τ_{eddy} and number density $l_{eddy}/\Delta x$ of eddy. This random field is generated in each sub-region computed in parallel during time integration of the model. The generation is based on a random seed given as a namelist parameter. α_{tuning} is a scaling factor to magnify the amplitude of the perturbations. However, the total amplitude is inherently scaled by the standard deviation of the tendency $\sqrt{\overline{\Phi'^2}}$. The length of eddies l_{eddy} is set to 1000 metres. The typical lifetime of convective eddies used for a temporal autoregressive process is 10 minutes. The scaling factor α is 5.0. For all simulation using the PSP scheme, the number of the IBC ensemble member is used as a random seed generating a stochastic pattern, i.e. the seed for ensemble member 1 is 1, and for the member 20 is 20 (as in Puh et al., 2023).

2.2.3 Microphysical parameter perturbations

To investigate the impact of microphysical uncertainty we use six combinations spanned by three different CCN concentrations and two different shape parameters of CDSD, similar to Matsunobu et al. (2022). These perturbations necessitate the use of the two-moment bulk-microphysics scheme (Seifert and Beheng, 2006), that is currently used in the pre-operational Rapid Update Cycle suite at DWD.

Perturbations in CCN concentrations consist of three pre-defined parameters: maritime (N_{CN} = 100 cm^{-3}), continental (N_{CN} = 1700 cm^{-3}), and polluted (N_{CN} = 3200 cm^{-3}) aerosol load based on Hande et al. (2016). 'Maritime' emulates clean, pristine conditions that have quite small numbers of CCN typical for the sea. 'Continental' is the default setting that represents the typical CCN concentrations for the European continental regions. The 'polluted' setting represents extremely polluted situations caused by, for example, massive wildfires and considerable anthropogenic emissions.

We also vary the shape parameters of CDSD estimation. The CDSD is approximated following the generalised gamma distribution

$$f(x) = Ax^{\nu} \exp\left(-\lambda x^{\mu}\right) \tag{2.2}$$

where A is dependent on the number density of hydrometeor particles and λ is a coefficient dependent on the average particle mass. The coefficients ν and μ are parameters that are

pre-defined and fixed throughout a simulation. In this study we control the widths of the CDSD by varying the shape parameter ν between 0 and 8 to cover a wide spectrum of the possible shape parameter values (Barthlott et al., 2022b). Since the parameters describing the CCN concentration and the shape of CDSD are kept temporally and spatially constant throughout a simulation, the MPP represent a structural uncertainty mimicking rather model error due to the incomplete knowledge of physical parameters than subgrid-scale variability.

2.3 Flow-dependent measure: convective adjustment time scale

Modulations of the convective environment are discriminated using the convective adjustment time scale τ_c (Done et al., 2006; Keil and Craig, 2011) describing a time scale over which CAPE is consumed by precipitation and convective equilibrium is established. This objective measure to classify convective weather situations is defined as CAPE (Jkg^{-1}) over its removal that is expressed by the precipitation rate P ($\mathrm{kgs}^{-1}\mathrm{m}^{-2}$):

$$\tau_c = \frac{CAPE}{dCAPE/dt} = \frac{1}{2}\frac{c_p \rho_0 T_0}{L_v g}\frac{CAPE}{P} \tag{2.3}$$

where c_p (specific heat capacity), ρ_0 (reference density), T_0 (reference temperature), L_v (latent heat of evaporation) and g (gravitational acceleration) are constants.

As an example, fingerprints of representative precipitation patterns during different convective forcing regimes are displayed in Fig. 4.1. During non-equilibrium conditions, there is no general uplift and CAPE can accumulate until local processes trigger convection. Being controlled by local factors, the resulting precipitation field typically has an intermittent spotty character (Fig. 4.1a). The area-averaged τ_c attains fairly large values, especially before the onset of convective precipitation around noon (Fig. 4.1c). On the other hand, during equilibrium, ascending motions driven by a large-scale flow cause widespread heavy rainfall (Fig. 4.1d). In such conditions CAPE generated by large-scale processes is immediately reduced by convective activity and τ_c usually attains small values of less than one hour (Fig. 4.1f).

2.4 Scale-dependent analysis: Fractions Skill Scores

Spatial error and spread

In this study, a variant of the Fractions skill score (Roberts and Lean, 2008) is used as a metric to examine the scale-dependent forecast error and spread of precipitation. The FSS is a fuzzy scoring technique quantifying the similarity between two binary fields (denoted A and B, observation and forecast fields in error, or two distinct ensemble members in spread) in terms of a predefined neighbourhood scale (Fig. 2.2).

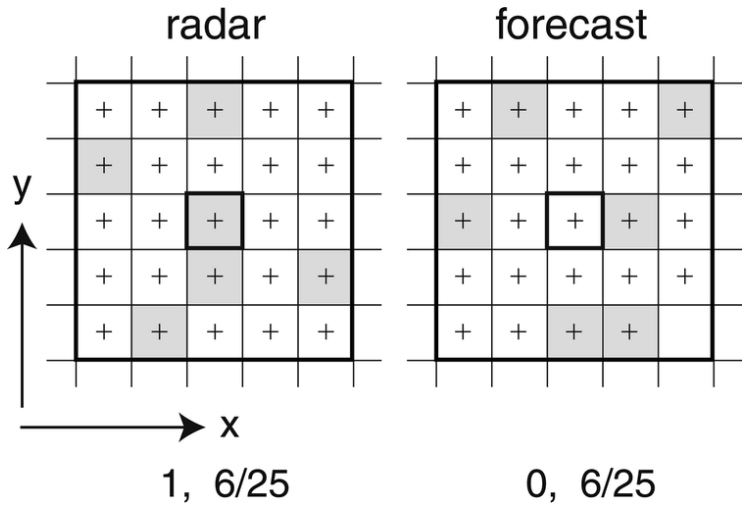

Figure 2.2: Binary field of precipitation after applying a threshold. Grey boxes show grid cells with precipitation above a certain threshold and white boxes illustrate non-precipitating grid cells or those with precipitation below the threshold. Two thick black boxes show two different neighbourhood sizes. The figure is adopted from Roberts and Lean (2008).

The definition of the FSS is given by

$$\mathrm{FSS} = 1 - \frac{\sum (f_A - f_B)^2}{\sum f_A^2 + \sum f_B^2} \tag{2.4}$$

where f_A and f_B represent the fraction of rainy grid points in fields A and B, respectively, at which the precipitation amount is above a certain threshold value. The second term on the right-hand side is the ratio of the mean squared error (MSE) of the fraction fields A and B to the maximum possible MSE. If the number of grid points with a value of 1 within a certain neighbourhood of a grid point is equal between two fields, the FSS is 1.0, which means the compared two fields are identical at the scale of the neighbourhood window. The FSS becomes smaller as the difference between two fields gets larger, and it becomes 0.0 when only one of the fields has values and the other has a complete miss in the respective neighbourhood.

It is known that the FSS is quite sensitive to the fraction of precipitating grid points in the entire field (Mittermaier and Roberts, 2010; Skok and Roberts, 2016, 2018). To remove the effect of frequency bias, we use the 95^{th} percentile values of hourly precipitation as the threshold to generate binary fields. The percentile threshold keeps the number of grid points used for the FSS calculation constant. Throughout the study, hourly forecast and observed precipitation rates are used for the evaluation. If the number of rainy grid points is less than 5% in the evaluation domain (Fig. 2.1), the field at that time is regarded as a complete miss. If both compared fields are a complete miss, that combination of fields is excluded from the analysis. The neighbourhood sizes are varied from 2.2 km (1 grid point) to 336.6 km (153 grid points) with an interval of 2 grid points. The largest window size

corresponds to half of the shorter side of the evaluation domain encompassing 294 × 341 grid points.

While the FSS was originally developed for comparing observations and forecasts, it can also show the dispersion of two fields (Dey et al., 2014). In this study, spatial error and spatial spread are defined based on Dey et al. (2014)'s method in order to evaluate the spatial error-spread relationship. To enable the FSS to be directly comparable like a classical error-spread analysis, spatial error and spread are defined as one minus the ensemble mean of each FSS:

$$\text{Spatial error} = 1 - \frac{1}{N}\sum_{i}^{N} \text{FSS}_{\text{of}_\text{i}} \tag{2.5}$$

$$\text{Spatial spread} = 1 - \frac{1}{N(N-1)}\sum_{i}^{N}\sum_{j\neq i}^{N} \text{FSS}_{\text{f}_\text{i}\text{f}_\text{j}} \tag{2.6}$$

where N is ensemble size and the subscript $_{\text{of}}$ and $_{\text{ff}}$ means that the FSS is calculated between observation vs forecast and forecast vs forecast, respectively. Following Dey et al. (2014), we calculate FSS_{ff} for all combinations of ensemble members belonging to an ensemble. For instance, FSS_{ff} for a 20-member ensemble can be calculated $20 \times 19/2 = 190$ times, and for a 6-member ensemble 15 times. Given the robustness of the ensemble mean FSS to ensemble size (Necker et al., 2024), we can compare mean FSS for ensembles of different sizes.

The error-spread relationship of the FSS was first illustrated by Zacharov and Rezacova (2009). Their result shows that FSS-based spread is smaller than FSS-based error for five case studies. However, their FSS-based spread might be underestimated because their ensemble spread calculation was centred on a single reference forecast. Here we apply the method of Dey et al. (2014) to mitigate this problem.

Scale detection using displacement scale

Although the FSS is a powerful measure for quantifying spatial dispersion of intermittent fields such as convective precipitation, it does not provide a direct measure in physical space. To identify scales where there is a certain degree of error or spread, we use a displacement scale (DS), defined as the smallest neighbourhood window size as

$$\text{DS} = \min(\text{FSS} \geq 0.5 + \frac{f_0}{2}) \tag{2.7}$$

where f_0 is the fraction of grid points considered in the FSS calculation (the 95^{th} percentile threshold gives $f_0 = 0.05$). Statistically, the DS is the smallest scale at which the forecast contains more useful information than a random forecast (Roberts and Lean, 2008). Half of the DS roughly corresponds to the distance of a displaced object between two compared fields with no frequency bias of binary fields (Mittermaier and Roberts, 2010; Skok and Roberts, 2018). Although the actual displacement length depends on many factors such as the sizes and shapes of an object (Skok and Roberts, 2018), we use the displacement

scale as a reasonable estimate of the doubled displacement length in this study. The DS can be considered directly as an estimate of displacement here because we use percentile thresholds to calculate the scale for each combination of field comparisons. Note that the DS is based on the same definition as the 'skillful' (Mittermaier and Roberts, 2010) and the 'believable' scale (Dey et al., 2014; Bachmann et al., 2020).

In the remainder of this paper, we refer to the displacement scale of spatial error and spread as eDS and sDS, respectively. The eDS is the smallest scale at which an ensemble shows a better prediction than a random prediction. The sDS is the largest scale at which significant spatial variability is achieved among ensemble members. If the FSS never exceeds $0.5 + \frac{f_0}{2}$, the displacement scale is set to be as large as the length of the evaluation domain. However, this no longer represents the scale of misforecasts – precipitation events are likely to be missed or false alarmed in this case. For this reason we examine the ensemble displacement scales using the median rather than the mean.

2.4.1 Verification dataset: Radar observations

An assessment of the practical predictability of convection, in particular its scale-aware component, requires a sound spatio-temporal dataset of precipitation measurements. The observations used for verification in this study are obtained from the German weather service (Deutscher Wetterdienst, DWD)'s observation network RADOLAN (Deutscher Wetterdienst (DWD), 2022). We used the RY product, whose valid observation domain covers most of the verification domain, with the missing domains along the south-western boundary (Fig. 4.1b,d). Near-surface reflectivities are observed every five minutes with 17 C-band radars at a spatial resolution of 1 km. The reflectivities are converted to five-minute precipitation rates and calibrated with ground ombrometer measurements. They are then accumulated to give hourly precipitation rates and regridded to the identical rotated pole grid as the ICON-D2 outputs for evaluation. Thus, the radar precipitation data comprise a synthesis of two data sources, radar and ground measurement network.

2.5 Variance-based analysis: variability budget

2.5.1 Sampling strategy

A sampling strategy basically follows the Morris method introduced in Sect 1.3. Morris' sampling strategy is a "One At a Time" method that iteratively perturb one parameter at a time. Figure 2.3a shows an example of the Morris' sampling strategy. Let us consider a model $f(p_A, p_B)$ with two uncertain input parameters p_A, p_B, assumed to be continuous and bounded between 0 and 1. Discretising the parameters into n_p levels creates an interval size of $\Delta = \frac{1}{n_p - 1}$. This results in p^2 possible combinations of p_A and p_B. The method begins by selecting an initial point (p_A, p_B) and performing a simulation. Next, a random parameter, say p_A, is perturbed by Δ. Following this, the remaining parameter, p_B, is also perturbed by Δ. This process results in three simulations, forming a trajectory that

captures the impact of parameter changes. Within the obtained trajectory, the changes in the model output due to individual parameter perturbations are termed elementary effects and defined as:

$$\begin{aligned} EE_A &= \frac{f(p_A + \Delta, p_B) - f(p_A, p_B)}{\Delta} \\ EE_B &= \frac{f(p_A + \Delta, p_B + \Delta) - f(p_A + \Delta, p_B)}{\Delta}. \end{aligned} \tag{2.8}$$

The mean and standard deviation of elementary effects are commonly used to assess sensitivity with the Morris sampling. Their stability with relatively small sample sizes reduces the influence of sampling uncertainty arising from the limited number of simulations used to calculate them. As each trajectory covers $\frac{1}{n_p - 1}$ of a parameter space, a minimal set requires covering all paths between discretised values with $r = (n_p - 1)$ repetitions. This translates to a minimum of $3(n_p - 1)$ simulations to generate a dataset suitable for deriving sensitivity indices.

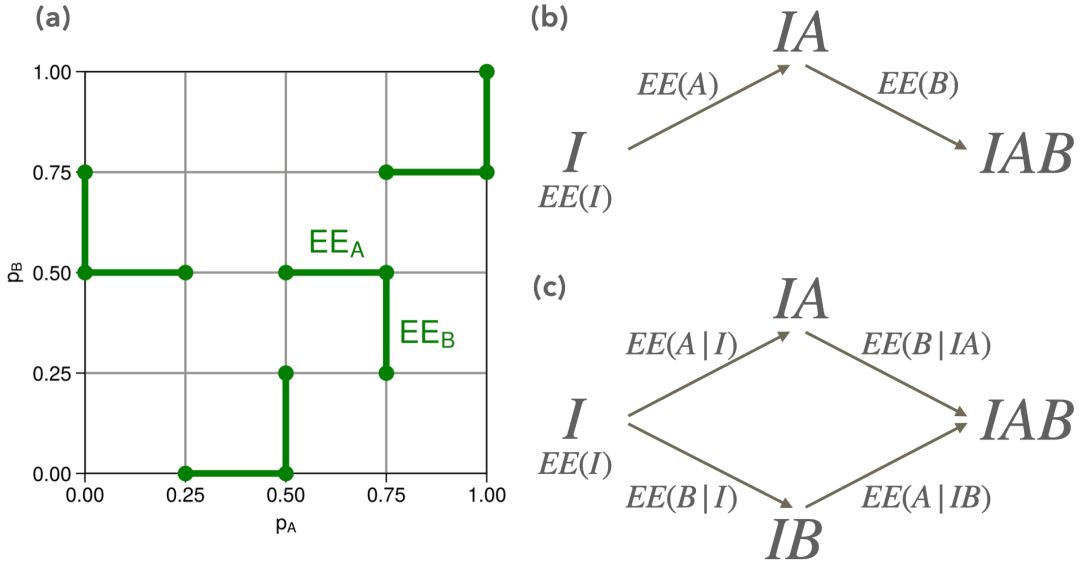

Figure 2.3: An example of Morris screening method (a) and pathways to add uncertainty sources (b, c).

We adopt a simplified strategy to construct an ensemble design suitable for investigating the impact of model uncertainty. The first simplification is to assume that $\Delta = 1$. Our focus lies on the variability introduced by adding a new model uncertainty representation, not on systematic biases that can be scaled by an amplitude of parameter perturbation. Therefore, binary perturbations ('on' or 'off') are sufficient. Second, instead of screening a starting point in the parameter space, we leverage IBC-perturbed members to define diverse starting points for each trajectory. While the parameter starting point remains fixed, the

varying atmospheric states introduced by different IBCs effectively sample the indirect effects of starting point variation. Third, the repetition r is chosen as a configurable sample size. Since binary perturbations necessitate a repetition r equal to 1 for each uncertainty representation, the minimum set of trajectories is 1 and does not capture the variability arising from different starting points. To imitate the variability due to the starting point and derive statistics of the elementary effects, we sample trajectories by perturbing IBCs. It is also possible to treat IBC perturbations as another uncertainty representation rather than a random initial state. However, since model uncertainty's impact depends on the input atmospheric state, the choice of the control IBC significantly constrains the location and intensity of model uncertainty's impact. It is more reasonable to perturb model-related inputs with IBCs to acquire correct understanding.

The above simplifications suggest a simple dataset design, consisting of a path of accumulated uncertainty representations (Fig. 2.3b). Let us consider a trajectory $f_i(x_A, x_B)$ initiated with the i-th IBC ($0 \leq i \leq r$) and subjected to two model uncertainty representations, A and B. The elementary effects within this i-th trajectory are defined as follows, adhering to the conventions of Eq. 2.8:

$$\begin{aligned} EE_I^i &= f_i(0,0) - \overline{f(0,0)}, \\ EE_A^i &= f_i(1,0) - f_i(0,0) \\ EE_B^i &= f_i(1,1) - f_i(1,0), \end{aligned} \tag{2.9}$$

where $\bar{\cdot}$ means an average over IBCs and EE_I is the IBC impact representing the variability of the starting points for model uncertainty perturbations. Additionally, we can calculate the mixed effect of multiple uncertainties, which is defined as a deviation from the ensemble mean when more than one uncertainty representation is turned on:

$$\begin{aligned} EE_{IA}^i &= f_i(1,0) - \overline{f(0,0)} \\ EE_{IAB}^i &= f_i(1,1) - \overline{f(0,0)}. \end{aligned} \tag{2.10}$$

These mixed effects represent the synergistic impacts resulting from perturbing multiple parameters simultaneously, and matches the sum of the individual elementary effects e.g. $EE_{IA}^i = EE_I^i + EE_A^i$. The ensemble statistics of these effects can be derived along the i axis with sample size r. We only use variance for the evaluation as it converges as quickly as standard deviation, which quickly converges with a small sample size compared to extreme quantile values (Craig et al., 2022). An investigation of variance convergence and sampling uncertainty is also performed in this thesis and the results are presented in Chapter 5.

Fédou and Rendas (2015) proposed a modified sampling strategy and another definition of the elementary effect. The proposed method is forming a cyclic graph with pathways for adding parameter perturbations to equitably consider the starting point dependence of impacts. The method maintains a fixed order of parameter perturbations within each trajectory. Then the sampling process is repeated with different orders until all possible order combinations are covered, forming a quadratic cycle graph like Fig. 2.3c. Under this

strategy, the elementary effects of p_A and p_B can be calculated as

$$\begin{aligned} EE_A &= \frac{f(p_A + \Delta, p_B) - f(p_A, p_B)}{\Delta} \\ EE_B &= \frac{f(p_A, p_B + \Delta) - f(p_A, p_B)}{\Delta}. \end{aligned} \tag{2.11}$$

For the two-parameter case, covering one pathway requires $2 \times n_p$ simulations, and a total of $2 \times (2 \times n_p)$ simulations are needed to cover all pathways. The difference between the mixed effect (Eq. 2.10) and the sum of elementary effects defined in Eq. 2.11 is considered as the higher-order mixed effect that represents higher-order non-linear interactions. Accounting for higher-order effects is crucial when the elementary effect depends on the sequence of perturbations (i.e. the starting point). For example, Covey et al. (2013) pointed out that the standard Morris elementary effect (Eq. 2.8) cannot differentiate impact arising from parameter interactions and its higher-order effects (interactions with pre-existing impacts). On the other hand, Awad et al. (2019) reported that the higher-order mixed effects can produce unrealistic rank of parameter importance in highly complex models. This is presumably due to insufficient sample size to estimate the higher order effect in such models. Therefore, in our study of a complex NWP model, we retain the definition in Eq. 2.8 but compare elementary and mixed effects across various pathways within the cyclic graph. This approach avoid potential shortcomings associated with relying solely on either approach. We can rewrite the elementary effect as a conditional elementary effect upon pre-existing perturbations. For instance in Fig. 2.3c:

$$\begin{aligned} EE^i_{B|IA} &= f_i(1,1) - f_i(1,0) \\ EE^i_{B|I} &= f_i(1,0) - f_i(0,0). \end{aligned} \tag{2.12}$$

The former is the elementary effect of parameter p_B within the upper pathway and the latter is within the lower pathway. Comparing $EE_{B|IA}$ and $EE_{B|A}$ allows us to understand how the elementary effect varies depending on a pathway. If they show similar statistics, uncertainty due to parameter p_B acts independently from pre-existing uncertainty. If not, the difference comes from the higher-order effects, meaning interactions with pre-existing uncertainty.

2.5.2 A decomposition of variance

Definition of variance budget equation

Utilising the elementary effects defined in the previous subsection, this section presents the rationale behind the budget decomposition of variance of the mixed effect. Let us consider an ensemble comprising two uncertainty representations, A and B. For simplicity their elementary effects defined as Eq. 2.8 are denoted as x_A and x_B and simply called (individual) impact of A and B. Similarly, the mixed effect defined as Eq. 2.10 is denoted as x_{AB} and called combined or synergistic impact. The corresponding bias-removed impacts

are x'_A, x'_B, and x'_{AB}, respectively. The variance of the combined impact (total variance), x'_{AB}, can be decomposed into the sum of the variances of the individual impacts, and their correlation term:

$$Var(x'_{AB}) = Var(x'_A) + Var(x'_B) + 2\sqrt{Var(x'_A)}\sqrt{Var(x'_B)}Corr(x'_A, x'_B). \tag{2.13}$$

This 'budget equation' reveals the efficiency with which uncertainty representations A and B contribute jointly to the total variance. A positive and large correlation coefficient $Corr(x'_A, x'_B)$ indicates that both representations perform similar tasks and enhance each other, potentially rendering one redundant. Conversely, negative correlation suggests that two representations cancel out each other. Correlation close to zero suggests orthognality of representations that effectively attain variability likely through distinct physical processes and the sum of individual variances becomes equal to the total variance in this case.

The budget decomposition also provides insights into factors influencing changes in total variance. Let us think of two cases where $Var(x'_{AB})$ has increased by adding uncertainty representation B to the pre-existing one A. If $Var(x'_A)$ or $Var(x'_B)$ is increased but $Corr(x'_A, x'_B)$ is unchanged, that signifies an increase in the variability of individual perturbations contributes to the total variance increase. Conversely, if $Var(x'_A)$ or $Var(x'_B)$ remain unchanged but $Corr(x'_A, x'_B)$ increases, the total variance increase stems solely from the modification in the correlation between x'_A and x'_B.

This decomposition of variance can be readily expanded to encompass n uncertainties:

$$Var(x_{total}) = \sum_i Var(x_i) + 2\sum_{i<j} \sqrt{Var(x'_i)}\sqrt{Var(x'_j)}Corr(x_i, x_j). \tag{2.14}$$

In this study, we applied this decomposition for the analysis of three uncertainties. The detailed calculation steps for this specific case are presented in Sect. 6.1.

Derivation

The derivation of the budget equation in Eq 2.13 proceeds as follows. We consider $\overrightarrow{x'_A}$ and $\overrightarrow{x'_B}$ as vectors in a 2D space, where their linear combination matches $\overrightarrow{x'_{AB}}$ with a certain correlation, as illustrated in Fig. 2.4a. Using the cosine theorem, we obtain the following identity:

$$|\overrightarrow{x'_{AB}}|^2 = |\overrightarrow{x'_A}|^2 + |\overrightarrow{x'_B}|^2 - 2\overrightarrow{x'_A} \circ \overrightarrow{x'_B}, \tag{2.15}$$

where $|\cdot|$ denotes the norm of a vector and $\circ$ denotes the inner product. Taking the mean over ensemble members yields:

$$\sigma(x'_{AB})^2 = \sigma(x'_A)^2 + \sigma(x'_B)^2 + 2Cov(x'_A, x'_B), \tag{2.16}$$

corresponding to a similar triangle, but with sides formed by the standard deviations $\sigma(x'_A)$, $\sigma(x'_B)$, and $\sigma(x'_{AB})$. Through a trigonometric identity, we derive the following:

$$\sigma(x'_{AB})^2 = \sigma(x'_A)^2 + \sigma(x'_B)^2 - 2\sigma(x'_A)\sigma(x'_B)\cos\theta, \tag{2.17}$$

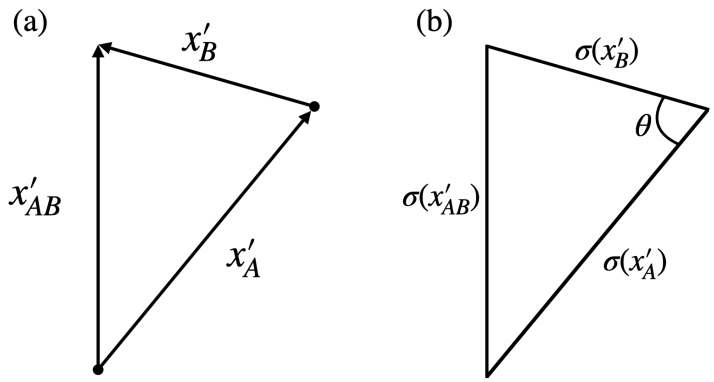

Figure 2.4: Trigonometric relationship of combined perturbations for (a) a single ensemble member and for (b) ensemble spread correlated with a coefficient satisfying $Corr(x'_A, x'_B) = -\cos\theta$.

where θ being the angle between the sides of x'_A, x'_B. Thus, the trigonometric relationship of a single member shown in Fig. 2.5a can be extended to that of ensemble spread (Fig. 2.5b).

Using the relationship $Corr(x'_A, x'_B) = \frac{Cov(x'_A, x'_B)}{\sigma(x'_A)\sigma(x'_B)}$, we can rewrite Eq. 2.16 using standard deviations:

$$\sigma(x'_{AB})^2 = \sigma(x'_A)^2 + \sigma(x'_B)^2 + 2\sigma(x'_A)\sigma(x'_B)Corr(x'_A, x'_B), \tag{2.18}$$

This is identical to Eq. 2.13. Comparing Eqs. 2.17 and 2.18, we obtain the following relationship:

$$Corr(x'_A, x'_B) = -\cos\theta. \tag{2.19}$$

When θ is a right angle $\pi/2$, the impacts of perturbing x'_A and x'_B are uncorrelated. When $\theta > 90$, the triangle becomes obtuse, indicating positive correlation. Conversely, $\theta < 90$ corresponds to an acute triangle and signifies negative correlation.

Figure 2.5 visually demonstrates the impact of correlation on total variance. In Fig. 2.5a, sides representing $\sigma(x'_A)$ and $\sigma(x'_B)$ intersect at an acute angle θ, indicating negative correlation. As the angle θ widens in Fig. 2.5b (gets more uncorrelated while side lengths remain the same), the total standard deviation $\sigma(x'_{AB})$ increases (grey solid and dashed circles). Thus, correlation serves as an indicator of how efficiently two impacts contribute to overall variability. When negatively correlated, the combination has small impact on forecast outcomes. Positive correlation suggests the perturbations are basically doing the same thing, implying that amplifying one perturbation might achieve similar results without additional computational cost.

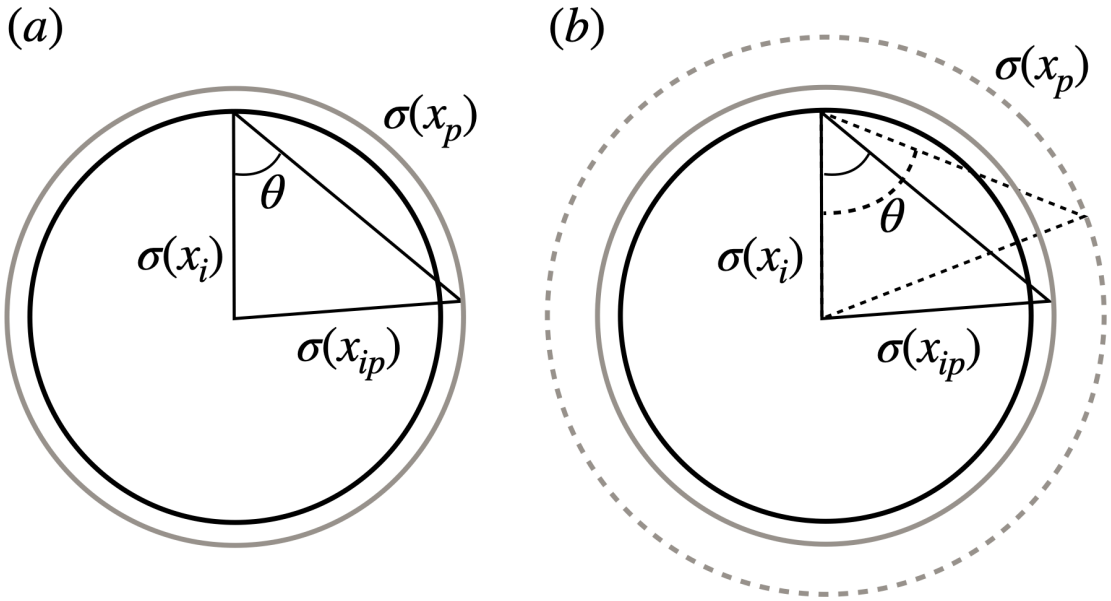

Figure 2.5: Trigonometric illustrations of ensemble spread increase. $\sigma(x'_A), \sigma(x'_B)$ are individual spread and $\sigma(x'_{AB})$ is spread of mixed effect. θ is a crossing angle satisfying $Corr(x'_A, x'_B) = -\cos\theta$. Black circles indicate ensemble spread purely due to x'_A. Grey circles indicate enlarged spread by adding x'_B. (a) a reference relationship of $\sigma(x'_A)$ and $\sigma(x'_B)$. (b) Further increased spread (grey dashed circle) when the correlation $-\cos\theta$ is weakened without a change in $\sigma(x'_A)$ and $\sigma(x'_B)$.

Comparison with Sobol's variance-based analysis

As briefly mentioned in Sect 1.3, the variance $Var(x'_{AB})$ can be decomposed into the sum of the Sobol's variances (Sobol', 1990):

$$Var(x'_{AB}) = V_A^S + V_B^S + V_{AB}^S, \tag{2.20}$$

where V_A^S and V_B^S are Sobol's first-order variances of x_A and x_B, defined as $Var(E[x'_{AB}|x'_A])$ and $Var(E[x'_{AB}|x'_B])$, respectively. V_{AB}^S is Sobol's interaction term. Under the condition $Var(x'_{AB}) \geq V_*^S$, the total variance can be described as the sum of non-negative Sobol's variances in this measure.

Using the law of total variance, the variance $Var(x'_{AB})$ can be written as a sum of expectation of the conditional variance and variance of conditional expectations on x'_A under the condition $Var(x'_{AB}) \geq Var(x'_A)$:

$$Var(x'_{AB}) = E_Z[Var(x'_{AB}|x'_A = Z)] + Var_Z(E[x'_{AB}|x'_A = Z]). \tag{2.21}$$

In this two-variable example, the first term on the right-hand side means mean variance given by x'_B while parameter A is fixed. Thus, the second term on the right-hand side represents the net gain in total variance due to adding uncertainty representation of A to

the pre-existing B. This term corresponds to Sobol's first-order variance:

$$V_A^S = Var[E(x'_{AB}|x'_A)] = Var(x'_{AB}) - E[Var(x'_{AB}|x'_A)]. \tag{2.22}$$

The Sobol's first-order variance is a residual of the subtraction of the mean variance due to all other impacts (only x'_B in this two-variable case) from the total variance. Thus, the following relationship stands out between the Sobol's index and the variability budget method:

$$\begin{aligned} V_A^S &= Var(x'_{AB}) - Var(x'_B) \\ &= Var(x'_A) + 2Cov(x'_A, x'_B). \end{aligned} \tag{2.23}$$

Consequently, the Sobol's interaction term can also be related to the terms in Eq. 2.16:

$$\begin{aligned} V_{AB}^S &= Var(x'_{AB}) - V_A^S - V_B^S \\ &= Var(x'_{AB}) - Var(x'_A) - Var(x'_B) - 4Cov(x'_A, x'_B) \\ &= -2Cov(x'_A, x'_B). \end{aligned} \tag{2.24}$$

In a two-variable case, both the Sobol's interaction term and the variability budget method's covariance term represent the same underlying interaction. We can obtain identical amount of the gain in the total variance from both methods.

One advantage of the variability budget method is that the individual variance (not the Sobol's first-order variance, which is conditional variance of ensemble mean) is already known and correlation coefficient can be derived without additional computations. Another advantage appears when the number of perturbed parameter become more than two, in which case the variability budget method still can decompose interaction terms into all pair-wise correlations, but the Sobol's method treats such interactions as a black box.

Chapter 3

Relative impact of MPP on daily precipitation and hydrometeors

The following Chapter is adapted from the publication titled "The impact of microphysical uncertainty conditional on initial and boundary condition uncertainty under varying synoptic control"(Matsunobu et al., 2022).

3.1 Background

Impact of uncertainties on convective-scale forecasts varies greatly depending on the prevailing flow conditions. A successful approach to classifying convective regimes is to focus on the strength of forcing that is driving convection (e.g. Keil et al., 2014, 2020; Flack et al., 2018). An objective measure of such a classification is the convective adjustment timescale (Done et al., 2006), which provides an indicator of the convective equilibrium measured as a timescale over which the convective available potential energy is consumed by convective precipitation.

In order to answer the first question posed in Chapter 1, **How does the relative impact of individual and combined uncertainties vary under different convective forcing regimes?**, this chapter assesses the impact of perturbations in IBC and microphysical parameters (MP) on convective precipitation and cloud formations on daily timescale. Five case studies are categorised into convective forcing regimes according to the timescale (refer to Sect. 2.3 for the detail method).

In Sect. 3.2, the dataset and approach are presented. Sect. 3.3 and 3.4 examines a case study for each forcing to quantify the relative impact on daily precipitation and cloud components, respectively. Sect. 3.5 aggregates results over case studies for systematic evaluation. Sect. 3.6 summarises the main results.

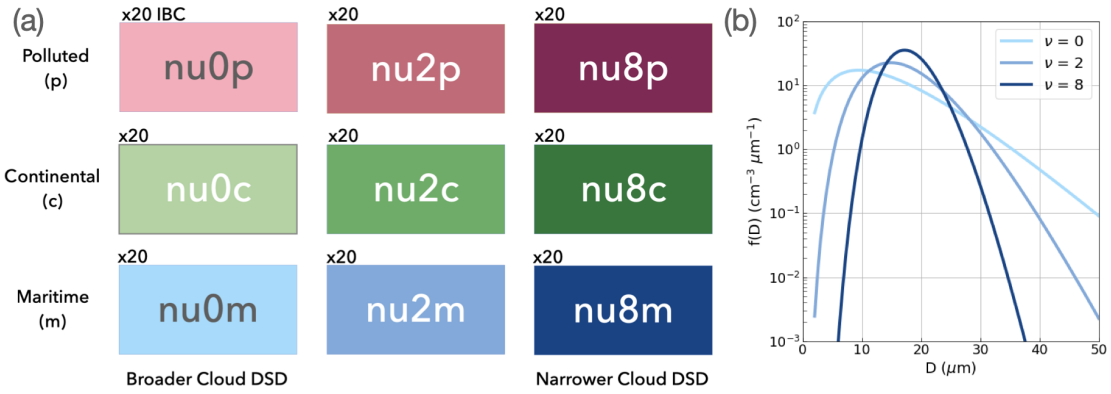

Figure 3.1: (a) Design of microphysically perturbed ensemble experiments. The colours used throughout the article indicate the nine different 20-member IBC sub-ensembles sharing the same combination of CCN and CDSD parameters. (b) Cloud droplet size distribution with different shape parameter ν at fixed cloud water content ($QC = 1\,\mathrm{g\,m^{-3}}$) and cloud droplet number concentration ($QNC = 300\,\mathrm{cm^{-3}}$). D denotes the diameter of the droplets.

3.2 Data and sub-sampling approach

3.2.1 Case Study dataset v1: August 2020

Ensemble design

To investigate the influence of uncertainties on CCN concentration and the shape of the CDSD in the presence of characteristic IBC uncertainty, we perform numerical experiments using 20 different IBCs, 3 different CCN concentrations and 3 different shape parameters of the CDSD yielding in total a 180-member ICON-D2 ensemble (Fig. 3.1).

The initial conditions are provided by pre-operational analyses produced by ICON-D2-KENDA (Kilometer-scale ENsemble Data Assimilation (Schraff et al., 2016). In August 2020 conventional measurements like radiosonde, aircraft and ground-based observations were assimilated in ICON-D2-KENDA using the local ensemble transform Kalman filter (LETKF; Hunt et al., 2007). ICON-D2-KENDA produces 40-member ensemble analyses, while the first 20 analyses are used as initial conditions for ICON-D2 ensemble forecasts (as in operations at DWD) with 24 h lead time due to limited computational resources. Lateral boundary conditions are based on ensemble ICON global and EU-nest simulations initialised 3 h before the initial time of the ICON-D2 ensemble experiments. The initial conditions for the global and EU-nest simulations are the operational analyses provided by DWD with a grid spacing of 40 km for the global domain and 20 km for the nested EU domain. Different from our ICON-D2 ensemble simulations the one-moment microphysics scheme and the convection parametrisation for deep and mid-level convection are active in the ICON global and EU-nest simulations. The lateral boundary conditions are updated hourly using data from the EU-nest forecasts at lead times from 3 to 27 h.

To examine the microphysical uncertainty we perturb the width of the CDSD and the amount of aerosol in the atmosphere by altering the CCN concentration. In the Seifert and Beheng (2006) scheme, CCN activation rates are calculated using a look-up table of activation rates empirically estimated by Segal and Khain (2006). To take insoluble CCN into account, certain portions of CCN are not activated depending on their particle sizes (Seifert et al., 2012). Consistent with Barthlott et al. (2022a, b) we vary CCN concentrations between pristine conditions and extremely polluted conditions. We employ three CCN concentrations: maritime (N_CN = 100 cm3), continental (N_CN = 1700 cm3) and polluted (N_CN = 3200 cm3). The "maritime"emulates clean, pristine conditions that have quite small numbers of CCN like over the sea. The "continental"is the default setting that mimics the observed CCN concentrations for the European continental regions (Hande et al., 2016). The "polluted"represents extremely polluted situations caused by, for example, massive wildfires and considerable anthropogenic emissions. The different CCN sub-ensembles that share the same CCN concentration are named with suffixes m(aritime), c(ontinental) and p(olluted), as shown in Fig. 3.1a.

The size distribution of hydrometeors is approximated using the following generalised gamma distribution:

$$f(x) = Ax^{\nu} \exp\left(-\lambda x^{\mu}\right) \tag{3.1}$$

where A is dependent on the number density of hydrometeor particles, and λ is a coefficient dependent on the average particle mass. The coefficients ν and μ are parameters that are pre-defined and fixed throughout a simulation. For example, with $\mu = \frac{1}{3}$ and $\nu = -\frac{2}{3}$, we can obtain the so-called Marshall–Palmer distribution of raindrops. In this study we control the widths of the particle size distributions by varying the shape parameter ν (for details see Barthlott et al., 2022a, b). With increasing ν the CDSD becomes narrower and more skewed as shown in Fig. 3.1b, which means the number concentrations of particles close to the mean size increase. In this study ν is varied between 0, 2 and 8 to cover a wide spectrum of the possible shape parameter values (as in Wellmann et al., 2020; Barthlott et al., 2022a, b). Note that the default setting is the broadest CDSD $\nu = 0$. Since the parameters describing the CCN concentration and the shape of the CDSD are kept temporally and spatially constant throughout the simulation, they rather represent model error due to the incomplete description of physical processes than subgrid-scale variability.

To address individual or combined impacts of forecast uncertainties mentioned above, we employ a sub-ensemble approach. A simple selection of different sub-ensembles sharing the same uncertainty allows us to quantify the relative impact of the various uncertainties. To focus on the combined impact of the microphysical perturbations, for instance, we can inspect 20 microphysical sub-ensembles consisting of nine members each sharing the same IBC but different combinations of CCN and CDSD parameters (microphysical (MP) sub-ensemble). To focus on the impact of the IBC uncertainties, we have nine IBC sub-ensembles available consisting of 20 members each (IBC sub-ensemble).

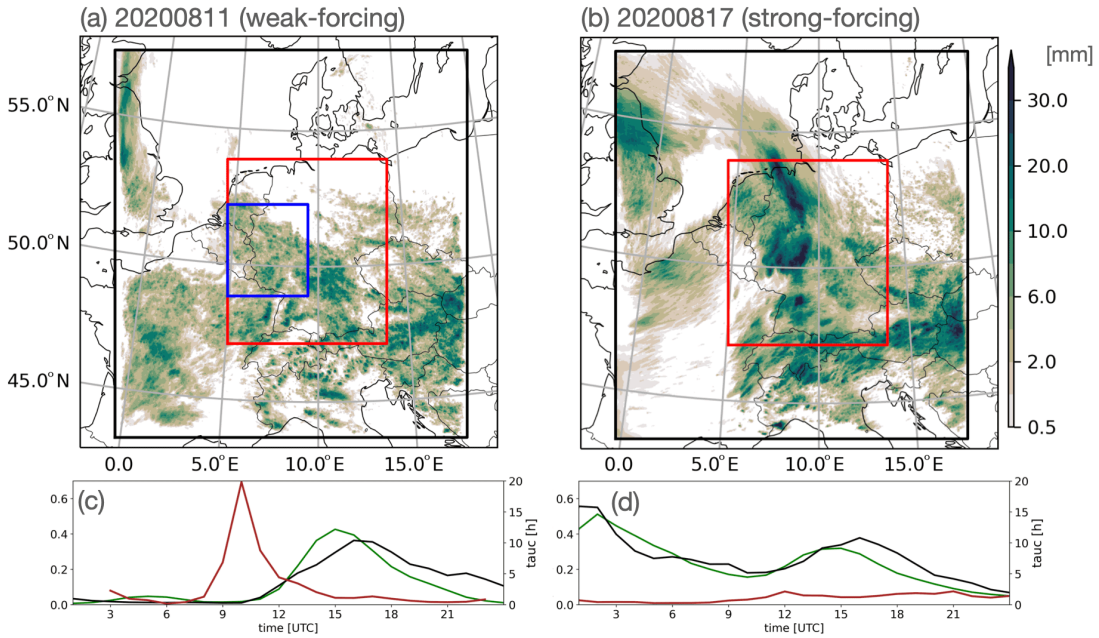

Figure 3.2: Daily accumulated precipitation on (a) a weakly forced day (11 August 2020) and (b) a strongly forced day (17 August 2020). Ensemble mean daily totals of the IBC sub-ensemble nu0c are shown. The black rectangles indicate the ICON-D2 simulation domain, the red rectangles depict the German domain used for evaluation, and the blue rectangle depicts the central western German domain used to inspect the spatial variability in rainfall patterns in Fig. 4.3 The time series of area-averaged hourly precipitation (green) and the convective adjustment timescale τ_c (red) complemented by the radar observed data (black) illustrate the different characteristics of both days in panels (c) and (d).

Weather situation and convective regime classifications

Two typical cases are selected for an in-depth investigation of the relative importance of the different uncertainties conditional on convective forcing. The distinct forcing situations are identified based on the convective adjustment time-scale τ_c (Sect. 2.3). On 11 August 2020, the precipitation texture shows a spotty distribution over southern Germany characteristic of convective precipitation in weak forcing situations (Fig. 3.2a). In a weak potential equivalent temperature gradient across central Europe (not shown), local trigger mechanisms (like convergence lines in the boundary layer caused by orography) initiate localised intense convection. The diurnal cycle illustrates the typical development of convective precipitation starting with little precipitation in the morning and peak precipitation in the afternoon (green line in Fig. 3.2c). The daily maximum value of τ_c peaks at about 20 h (red line in Fig. 3.2c), exceeding the 6 hour threshold used in previous work to distinguish different synoptic control in Europe (Keil et al., 2014, 2019; Kühnlein et al., 2014; Baur et al., 2018; Flack et al., 2018).

The 17 August 2020 represents a strong forcing situation associated with a weak low-

Table 3.1: List of case studies for which 180-member ICON-D2 ensemble experiments were performed, indicating the date, the type of synoptic forcing, the daily maximum convective adjustment time-scale (τ_c) and daily precipitation of different IBC sub-ensemble means with their microphysical configurations, respectively.

Date	Forcing	τ_c [h]	Precipitation [mm/d]		
			default	maximum	minimum
11 August 2020	weak	20	2.67 ($nu0c$)	2.95 ($nu8m$)	2.42 ($nu8p$)
12 August 2020	weak	7	1.58 ($nu0c$)	1.73 ($nu8m$)	1.45 ($nu8p$)
13 August 2020	strong	3	3.72 ($nu0c$)	3.90 ($nu8m$)	3.60 ($nu2p$)
17 August 2020	strong	2	5.72 ($nu0c$)	6.00 ($nu8m$)	5.51 ($nu8p$)
18 August 2020	weak	6	3.79 ($nu0c$)	4.07 ($nu0m$)	3.51 ($nu8p$)

pressure system located over France that moved eastward towards Germany (not shown). The cyclonic flow favoured large-scale ascent initiating convection, especially over the western part of Germany, resulting in widespread precipitation (Fig. 3.2b). There was rainfall from the start of the forecast, and the heaviest rainfall occurred at night followed by a gradual reduction in precipitation until noon (green in Fig. 3.2d). In the afternoon, there was a secondary peak of convective precipitation between 11:00 and 18:00 UTC. The daily maximum τ_c is less than 2 h on 17 August 2020 (Table 3.1, and red line in Fig. 3.2d). Such low values indicate that CAPE was immediately consumed by a continuous triggering of convection caused by synoptically forced ascending motion characteristics in a so-called equilibrium regime.

The comparison of the precipitation time series with area-averaged radar observations indicates the realism and fidelity of the ICON-D2 ensemble forecasts (Fig. 3.2c and d). Characteristic values of the remaining three cases and their classification are presented in Table 3.1.

3.2.2 Sub-sampling approach

We extract different sub-ensembles from the large 180-member ensemble. First we create nine-member MP sub-ensembles in which each of the sub-ensemble members has different combinations of CCN and CDSD parameters but identical IBCs to examine the relative contribution of the combined MP perturbations on precipitation. Since there are 20 IBCs in the entire ensemble, there are 20 different MP sub-ensembles with nine members each. Likewise, we can make nine 20-member IBC sub-ensembles as well, with one fixed combination of MP parameters but 20 different IBCs. These different sub-sampling perspectives allow conclusions to be drawn about the relative impact compared to IBC uncertainty. Last, we also make 60 three-member CCN and CDSD sub-ensembles to further decompose the impact into the contributions of each.

3.3 Systematic impact on daily precipitation

To assess the relative contributions of the various uncertainties we extract different sub-ensembles from the large 180-member ensemble. First we focus on 9-member MP sub-ensembles in which each of the sub-ensemble members has different combinations of CCN and CDSD parameters but identical IBC to examine the relative contribution of the combined microphysical (MP) perturbations on precipitation. Since there are 20 IBC in the entire ensemble, there are 20 different MP sub-ensembles with nine members each. Likewise, there are nine 20-member IBC sub-ensembles, with one fixed combination of MP perturbations but 20 different IBC. This different subsampling perspective allows drawing conclusions on the relative impact of IBC uncertainty. Lastly, there are sixty 3-member CCN and CDSD sub-ensembles, respectively, that inform on their individual contribution.

To estimate the impact of the combined microphysical uncertainty we first focus on nine-member microphysics (MP) sub-ensembles sub-sampled from the entire 180-member ensemble. The relative differences in 24 h accumulated area-averaged precipitation forecast of all 180 ensemble members to their combined MP sub-ensemble mean are shown in Fig. 3.3 for a synoptically weak and a strong forcing case to contrast the flow-dependent behaviour. Every dot represents the precipitation difference in a single ICON-D2 forecast to its sub-ensemble mean. Since there are 20 different MP sub-ensembles composed of nine microphysically perturbed members (colour coded as in Fig. 3.3), the 180 dots illustrate the overall variability. Apparently the impact of microphysical uncertainty is larger during weakly forced conditions, and there is surprisingly high variability between the different MP sub-ensembles, in particular during weak control. The largest and smallest range of precipitation differences amounts to 48 % (+23 % to 25 %) and 11 % (+7 % to 4 %), respectively (compare members eight and nine in Fig. 3.3). During strong synoptic control the differences amount to 16 % (+9 % to 7 %) and 4 % (+2 % to 2 %), respectively (compare members 2 and 18 in Fig. 3.3).

Furthermore, it is possible to assess the different microphysical impact on precipitation. The average precipitation differences caused by MP perturbations are displayed by coloured lines in Fig. 3.3; for instance, experiment nu8m (narrow CDSD and maritime CCN content; dark blue) exhibits the largest precipitation deviations in both regimes. More generally, experiments with maritime aerosol load (low CCN content, blue) show an increase in precipitation, while the experiments with high CCN concentrations (polluted, red) show a decrease. Increasing the CCN concentration from maritime (nu8m) to polluted conditions with narrow CDSD shape (nu8p) amplifies average precipitation differences to +11 % and 14 % in the weak forcing case, respectively (+5 % to 4 % in the strong forcing case). A comparison between the lines having the same colours but a different darkness shows that the shape parameter of the CDSD also exhibits a systematic impact in the weak forcing situation (e.g. light red (nu0p) and dark red (nu8p) lines in Fig. 3.3a), whereas a CDSD's impact is hardly seen in the strong forcing situation. Narrower CDSD distributions give less precipitation, particularly during polluted conditions (nu8p, dark red). The larger sensitivity to CDSD during weak synoptic control and a systematic decrease in precipitation with increasing shape parameters are consistent with Barthlott et al. (2022a, b). During

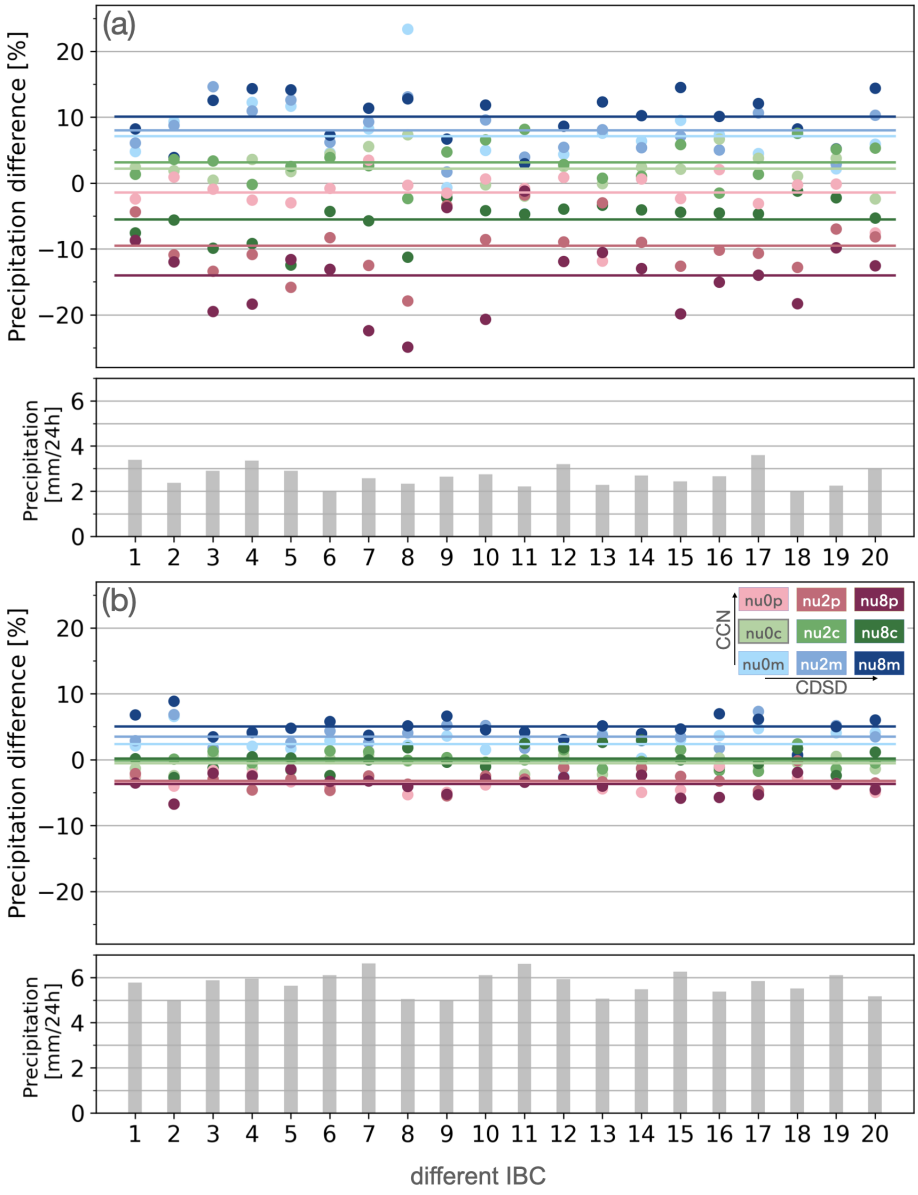

Figure 3.3: Relative difference of daily area-averaged precipitation [%] with respect to combined microphysical (MP) sub-ensemble means sharing the same initial and lateral boundary conditions (IBC) for the (a) weak and (b) strong forcing case. The columns below indicate absolute precipitation values of the 20 different MP sub-ensemble means. The nine colours indicate all combinations of microphysical configurations (as in Fig. 3.1a). The coloured lines show the averages.

strong synoptic control the average relative difference is governed by the CCN concentration (Fig. 3.3b).

The governing role of IBC perturbations on precipitation is evident when comparing the sub-ensemble mean precipitation amounts of the 20 different MP sub-ensembles. During weak control, the variability ranges between 1.9 and 3.6 mm d^1, whereas it ranges between 5.0 and 6.6 mm d^1 during strong synoptic control (lower panels in Fig. 3.3). This variability is purely caused by IBC uncertainty driving the 20 different MP sub-ensembles. The similar amplitude of the variability (1.7 versus 1.6 mm d^1) suggests a larger impact of IBC uncertainty during weak control when the absolute rainfall values are roughly only half as large. There is no systematic relationship between the precipitation amount and the amplitude of relative differences during both regimes. That means the microphysical impact is not constrained by daily precipitation amounts.

Interestingly, a closer inspection reveals that different IBCs can completely reshuffle the rank of the individual members in a specific MP sub-ensemble. For instance, experiments with modest aerosol content but different shapes of the CDSD show extremes for member 11 during weak control (nu8c (dark green) shows the largest negative and nu2c (medium green) shows the largest positive impact; Fig. 3.3a). This non-systematic and highly varying response of precipitation to perturbed microphysical parameters of individual ICON-D2 experiments points towards a strong sensitivity to IBC. This finding illustrates the necessity to be cautious when interpreting results based on a deterministic approach only to evaluate uncertainty.

Next, we further compress the data to directly compare and quantify the relative contribution of the various sources of uncertainty conditional on the weather regime. The resulting relative daily area-averaged precipitation differences of various sub-sampling strategies are displayed in Fig. 3.4. We again calculated the deviations with respect to a sub-ensemble mean; for instance, the nine different 20-member IBC sub-ensembles are shown by orange box and whisker diagrams depicting the medians, interquartile ranges, 5th and 95th percentiles, and outliers.

First, it becomes evident that the magnitude of the impact of the various uncertainties largely depends on the synoptic control. The IBC sub-ensembles show a remarkable range of +38 % to 30 % in daily precipitation sums during the weak forcing situation (filled orange dots of IBC in Fig. 3.4). Although their medians and interquartile ranges have some variability among the different microphysics configurations, no systematic dependence is found, and the variability between the nine IBC sub-ensembles is statistically insignificant. A corresponding behaviour is found for the strong forcing case with smaller amplitudes between +15 % and 12 % (open orange dots in Fig. 3.4).

Secondly, the synergistic effect of microphysical perturbations (grey in Fig. 3.4) ranges between +22 % and −25 % for the weak forcing case, and ±10% for the strong forcing case. Note that the relative differencea of the 20 different MP sub-ensembles (with nine members each), previously discussed in detail (Fig. 3.3), are collapsed into one column here.

The individual microphysical perturbations consequently result in 60 sub-ensembles (with three members each) denoted CCN sub-ensemble and CDSD sub-ensemble. Inter-

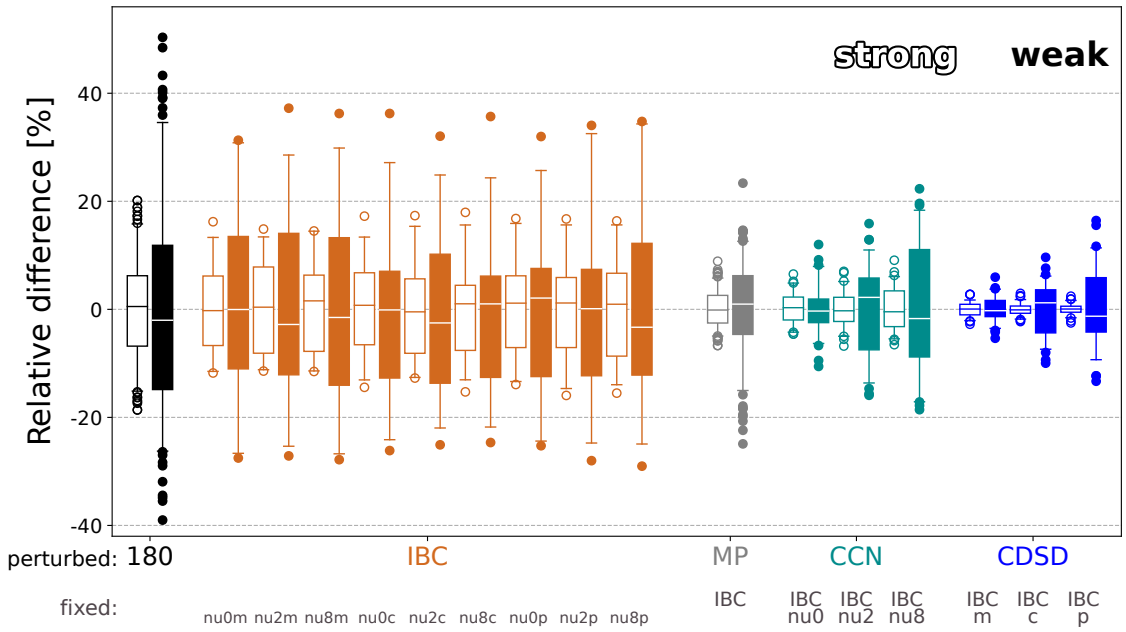

Figure 3.4: Box and whisker diagram showing the relative differences of daily area-averaged precipitation of individual ICON-D2 members belonging to various (sub-)ensembles. The perturbations (x labels in colour) and different fixed configurations (grey x labels) are indicated. 180 is the abbreviation of the entire ensemble, IBC, MP, CCN and CDSD for the different sub-ensembles. The bars, boxes, whiskers and dots show medians, interquartile ranges, 5th and 95th percentiles and outliers, respectively. Filled boxes represent weak control (11 August) and open boxes strong synoptic control (17 August).

estingly, the impact of individual CCN perturbations shows a clear dependence on the CDSD shape and vice versa. The CCN impact is smallest (±10%) with a broad distribution (shape parameter $\nu = 0$) and increases to a range of +22 % and −20 % with narrower distributions (increase in shape parameter). The impact of CDSD perturbations also increases with an increase in the CCN concentration. This steady increase in impact is also found in the CCN concentrations during strong forcing, while the shape of the CDSD shows a small sensitivity only. Precipitation reacts more sensitive to microphysical perturbations during weak synoptic control. In this situation, the interquartile range of the combined MP sub-ensemble (grey box) becomes smaller than those of the CCN sub-ensembles with fixed shape parameters (cyan boxes for fixed $\nu = 2$ and 8) corresponding to a narrower CDSD. Thus adding CDSD perturbations to CCN uncertainty renders the probability density function of the relative impact sharper and leads to an extension of the tails of the distribution (grey dots of MP sub-ensemble).

Finally, the 180-member ensemble including IBC and microphysical uncertainty shows the largest variability during weak control. Conditional on the weather regime the extremes in daily precipitation of individual members deviate from the ensemble mean by +50 % to 40 % with an interquartile range of ±15 %. Interestingly the interquartile range and the

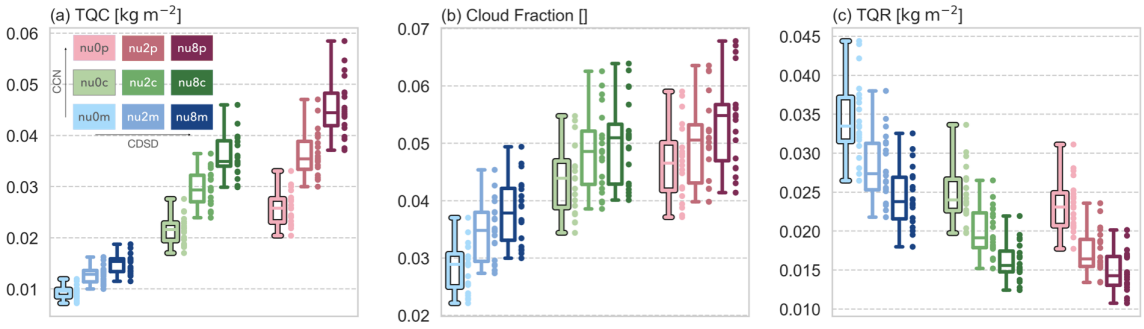

Figure 3.5: Box and swarm plots for 24h-mean (a) domain-averaged total column cloud water content, (b) cloud fraction and (c) domain-averaged total column rain water content over Germany for the weak forcing case. The boxplots and dots illustrate the same data set, but the dots represent individual IBC sub-ensemble members. The colours are based on the various combination of microphysical perturbations shown in Fig. 3.1a. Boxplots show medians, interquartile ranges as well as maximum and minimum values, respectively.

5th and 95th percentiles of the 180-member ensemble are similar to pure IBC uncertainty (compare black and orange box and whiskers). Again, microphysical uncertainty particularly affects the tails of the distribution (which are 10 % of the members represented as dots in Fig. 3.4).

In summary, IBC uncertainties dominate the impact on precipitation, while microphysical uncertainties play a secondary role. CCN has a larger impact than CDSD. Combined perturbations of CCN and CDSD enhance each other and show larger extremes in precipitation than individual CCN and CDSD perturbations. While the interquartile range of the 180-member ensemble and the individual IBC sub-ensembles is similar, the extreme members in the full ensemble surpass the IBC variability by +15% and -10%. Thus, the combination of IBC and microphysical uncertainty affects the magnitude of the extremes while keeping the interquartile range fairly unaffected.

3.4 Systematic impact on cloud and rainwater content

To complement the assessment centred on the relative impact on precipitation, we now turn to important precursors in the complex process chain to form precipitation and inspect the contribution of the uncertainties to the cloud and rainwater content within a full convective-scale EPS framework. Since we find similar systematic responses in both weather situations, we show results for the weakly forced case only. In Fig. 3.5 we depict the variability caused by IBC uncertainty in clouds and rainwater. The 24 h mean of hourly values is computed for the nine different IBC sub-ensembles to examine the relative impact.

The vertically integrated cloud water content (TQC) increases significantly with higher CCN concentration and CDSD shape (Fig. 3.5a). The medians of the ensembles with different microphysics uncertainty vary by more than 400 % (TQC is amounting to 0.01 kg m2 in experiment nu0m and 0.044 kg m2 in nu8p). The comparison of sub-ensembles

sharing identical CDSD shape parameters shows an increase in TQC of up to 300 % when increasing CCN concentrations from maritime to polluted conditions (compare experiments nu0m and nu0p in Fig. 3.5a). Similarly, the change from the broadest to the narrowest CDSD enhances TQC by roughly 150 %. These values are more than an order of magnitude larger compared to the impact of microphysical perturbations on precipitation (compare to orange IBC sub-ensembles in Fig. 3.4). An important implication seen in Fig. 3.5a is that IBC perturbations cannot encompass the variability caused by microphysical uncertainties in cloud forecasts, which manifests by marginal (or no) overlap of the distributions which have different CCN and CDSD configurations (differently colour-coded in Fig. 3.5).

The forecast cloud fraction also systematically increases with higher CCN and shape parameters (Fig. 3.5b), in agreement with the increase in TQC. Cloudy grid points are defined as grid cells where TQC $>$ 50 g m2. The medians of the cloud fraction in IBC sub-ensemble nu0m (light blue), nu8m (dark blue), nu0p (light red) and nu8p (dark red) are 0.29, 0.39, 0.47 and 0.55, respectively. Thus, cloud fraction increases with higher CCN and/or CDSD parameters by 35 %, 62 % and 91 % relative to experiment nu0m. Compared to TQC, a change in CDSD shape parameters shows an only minor effect on cloud fraction in continental and polluted CCN conditions (e.g. nu8c and nu8p in Fig. 3.5b). This is presumably caused by ambient atmospheric conditions as, for example, humidity sets an upper bound for total cloud cover. Hence microphysical uncertainty (CCN and CDSD perturbations) becomes less important, and IBC uncertainty, which predominantly triggers convection and determines the upper bound of cloud coverage, governs the variability in spatial cloud distributions.

Finally, the vertically integrated rainwater content (TQR) averaged over Germany shows a systematic but opposite response compared to TQC (Fig. 3.5c). TQR decreases with increasing CCN and shape parameters of the CDSD and parallels the systematic impact found for precipitation. Compared to TQC the variability caused by microphysical perturbations becomes smaller; for instance, the TQR median of experiment nu0m amounts to 0.033 kg m2 and nu8p to 0.014 kg m2, indicating a decrease by roughly 58 %.

The steady decreasing systematic impact of the microphysical uncertainty on cloud water content, rainwater content and eventually precipitation hints towards some kind of buffering effects or compensating processes that reduce the large, positive impact on clouds and eventually even turn it into a negative impact with respect to rain production. Companion work by Barthlott et al. (2022a, b) and Baur et al. (2022) shed light on those processes. One major process is the reduction in warm rain processes. The suppression of collisional growth of cloud droplets in polluted CCN conditions reduces the formation of raindrops, and small droplets become more likely to evaporate. Moreover, cloud optical properties are influenced as well through changes in the droplet's effective radius. That, in turn, can affect the radiative energy supply that triggers new convection.

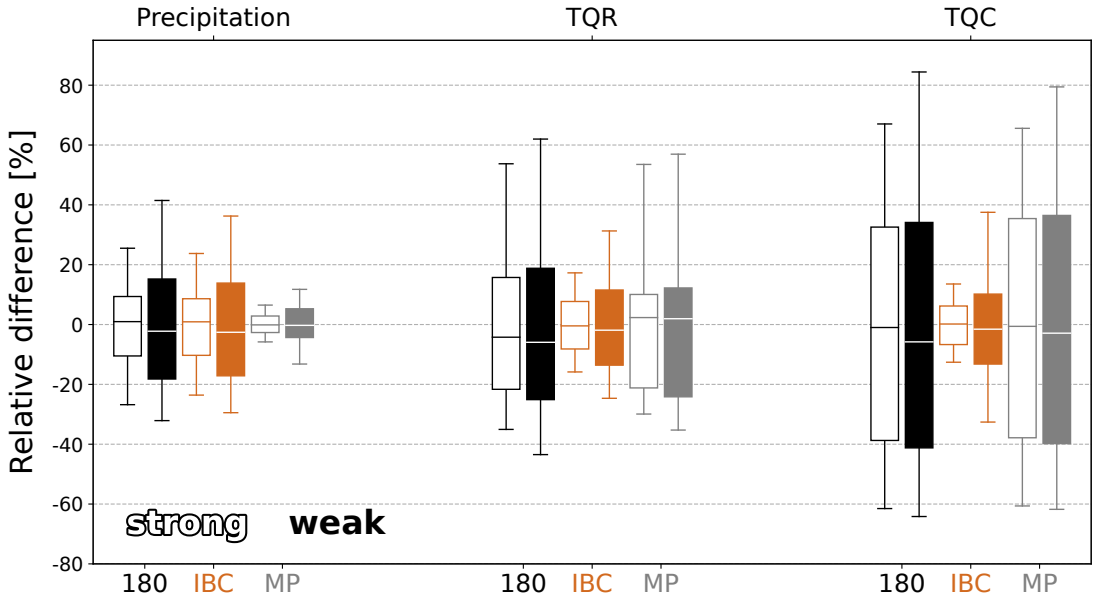

Figure 3.6: Relative differences of the 180-member ensemble (black), the averaged IBC sub-ensembles (orange) and averaged combined microphysical sub-ensembles (grey) aggregated over five days in August 2020 conditional on synoptic control. Relative differences of precipitation, total column rain water content (TQR) and total column cloud water content (TQC) are displayed using filled boxes for weak forcing situations. Boxplots show bootstrapped medians, interquartile ranges as well as the 5th and 95th percentiles, respectively. For details see the text.

3.5 Flow-dependent impact aggregated over cases

Finally, we repeat the analysis and use 180-member ICON-D2 ensemble experiments performed for 5 d in August 2020 to confirm the previous findings. The classification into distinct weather situations with different synoptic control results in three weakly and two strongly forced cases (see Table 3.1). The regime-dependent relative impact of the various perturbations is computed as follows: first, the relative difference in every individual member to its corresponding sub-ensemble mean is calculated separately for every day (as in Sect. 3.3 and displayed in Fig. 3.4). Secondly, the median, the interquartile range, and the 5th and 95th percentiles are computed by aggregating the days for each synoptic forcing separately (i.e. 540 samples for weak and 360 for strong forcing). Finally, the samples are bootstrapped 100 times with replacement to get robust results, and the mean of the 100 medians, interquartile ranges and percentile values are finally depicted in Fig. 3.6. This procedure takes into account the different mean values of distinct sub-ensembles on different days (see Table 3.1) and allows a fair comparison. The 5th and 95th percentiles of the relative difference then define the 90 % confidence interval (similar to Craig et al. (2022)).

In the full 180-member ensemble, including IBC and combined microphysical uncer-

tainties, the confidence interval of precipitation deviates for individual experiments from the ensemble mean by +41 % to 32 %, with an interquartile range between +15 % to 18 % during weak forcing. The corresponding impact of pure IBC perturbations shows a range of +36 % to 29 % during weak forcing (orange boxes of IBC sub-ensemble in Fig. 3.6). The variability is smaller and amounts to ±23 % during strong forcing. The medians have a slightly negative bias for the weak forcing cases because the precipitation distribution is slightly positively skewed; i.e. the mean is larger than the median. That might be an artefact of the given sample size.

The impact of combined microphysical perturbations on the confidence interval of precipitation (grey bars in Fig. 3.6) varies between ±12 % and 13 % in weak forcing cases and ±6 % during strong forcing cases. Thus precipitation amounts are twice as sensitive to pure microphysical perturbations during weak control. Adding microphysical perturbations to the IBC sub-ensembles (giving the full 180-member ensemble) shows a negligible impact on the interquartile range (compare black and orange bars in Fig. 3.6) but extends the tails of the distribution (black and orange whiskers in Fig. 3.6) by 5 % for weak forcing conditions.

The same methodology is applied to convective clouds represented by averaged vertically integrated rainwater content (TQR in Fig. 3.6). Microphysical perturbations show a larger impact than IBC perturbations. The confidence interval of the impact of microphysical perturbations on TQR ranges between ±54 % and 30 % for strong forcing and between +57 % and 35 % for weak forcing. Forecast variability is increased by +31 % compared to the pure IBC impact when taking the microphysical uncertainties into account, too. The relative impact of IBC perturbations on TQR ranges between +31 % and 25 % for weak forcing and between +17 % and 16 % for strong forcing.

Finally, the impact on vertically integrated cloud water content (TQC in Fig. 3.6) shows less dependence on synoptic control than those on rainwater or precipitation. Microphysical perturbations show a large impact on TQC, and their impact exceeds the impact of IBC uncertainty. The relative impact of microphysical perturbations on TQC ranges between +80 % and 62 % for weak forcing and between +66 % and 60 % for strong forcing. Forecast variability is increased by +47 % compared to the pure IBC impact when taking the microphysical uncertainties into account. The variability in CCN and CDSD plays a larger role in narrower CDSD or higher CCN conditions (not shown), similar to the impact on precipitation discussed in Fig. 3.4.

Overall, microphysical uncertainty plays a more important role in the prediction of cloud and rainwater content than IBC uncertainty, but the impact is buffered during warm rain processes. The buffering effect that counteracts microphysical perturbations discussed in the previous sections can thus be quantified. The microphysical impact on the 95th percentile value decreases from +79 % for TQC to +57 % for TQR and to +12 % for precipitation. We find a systematically larger impact of the various uncertainties for precipitation, TQR and TQC during weak forcing conditions.

3.6 Summary and Conclusions

The relative importance of microphysical uncertainties on cloud and precipitation forecasts in a full convective-scale EPS framework is assessed on different spatial and temporal scales conditional on synoptic control in central Europe. In the present study we perturb two microphysical parameters that are poorly constrained by observations. Those constitute the cloud condensation nuclei (CCN) concentration and the shape parameter of the cloud drop size distribution (CDSD), both currently not perturbed in operational ICON-D2 ensemble forecasts. An examination of the synergistic effect of these microphysical perturbations necessitates the use of the two-moment bulk microphysics scheme of Seifert and Beheng (2006) that predicts next to the mass concentration of different hydrometeors their number density and thus allows the calculation of the particle size distribution. Their individual and combined relative impact is estimated in the presence of initial and boundary condition uncertainty (IBC) available from operational ensemble forecasting at Deutscher Wetterdienst. Nine different set-ups of such combined microphysical perturbations run with 20 different IBC add up to 180-member ICON-D2 ensemble forecasts. The relative impact of the various uncertainties is quantified by selecting different sub-ensembles that are sharing a common uncertainty.

Based on five real summertime cases we find that the impact of the various uncertainties on precipitation crucially depends on the synoptic control. It is larger during weakly forced situations. The IBC uncertainty accounts for most of the precipitation variability. The 90% confidence interval (that is given by the 5th and 95th percentile) of daily area-averaged precipitation of individual ICON-D2 experiments ranges between +38% and -32% during weak forcing and ±23% during strong forcing (Fig. 3.6). Combined microphysical perturbations show a relative impact on the 90% confidence interval of precipitation between +12% and -13% during weak forcing, and ±6% during strong synoptic control. Thus precipitation amounts are twice as sensitive to microphysical perturbations during weak control. The joint effect of IBC and microphysical uncertainty extends the tails of the forecast distribution by 5% in weakly forced conditions. Individual ICON-D2 members exceed the ensemble mean precipitation by 50%. However, the interquartile range of the full ensemble only marginally deviates from the pure IBC sub-ensembles (Fig. 3.4).

Clouds react differently on the various uncertainties. The combined microphysical perturbations largely determine the variability of daily- and area-averaged vertically integrated cloud water content (TQC in Fig. 3.6). Different from their impact on precipitation, the increase of CCN concentration and shape parameter of CDSD has a large positive impact on the production of cloud and rain water content forming horizontally larger clouds. Further, this impact is fairly weather regime independent. Thus the considerable impact on cloud variables does not directly translate into precipitation amounts. This suggests that there are some microphysical processes or feedback mechanisms involved that compensate and ultimately reverse the impact of microphysical perturbations on clouds and precipitation. The systematic behaviour of cloud variables is consistent with previous studies (Seifert et al., 2012; Igel and van den Heever, 2017a; Wellmann et al., 2020; Zhang et al., 2021), and further discussion about the detailed processes seen from the deterministic per-

spective can be found in Barthlott et al. (2022b) and Baur et al. (2022). Note that we compare rainfall accumulations at the ground with averages of 24 hourly snapshot scenes of vertically integrated cloud and rain water to facilitate a comparison of the respective contribution.

Importantly, a close inspection of the impact of microphysical uncertainties in the presence of different IBC on precipitation indicates a strong sensitivity to IBC uncertainty (Fig. 3.3). This illustrates the necessity to be cautious when interpreting results based on a deterministic approach only to evaluate impact of uncertainty. The use of a full ensemble modelling framework including various key sources of uncertainty as done in this study is essential to assess their relative importance. Another major conclusion is the necessity to take the atmospheric state into account when quantifying the contribution of various uncertainties. Given that roughly 20 to 40% of the days with summertime precipitation in central Europe are classified as being weakly controlled (Kühnlein et al., 2014; Zimmer et al., 2011), the considerable impact during these conditions is usually veiled when inspecting results independent of the synoptic control. A limitation of this study is the limited dataset covering five days in August 2020 only. More robust results require a larger database containing more cases that comprise different synoptic conditions. Based on the five cases we cannot draw general conclusions. However, we believe that the findings are robust enough to provide a scientific basis for future research.

Our results suggest that the consideration of CCN and CDSD uncertainties increases precipitation variability and can contribute to the reduction of the long-standing issue of underdispersion of near surface variables in convective-scale EPS forecasts (see references in e.g., Keil et al., 2019) and thus ultimately benefit the improvement of NWP ensemble forecasting. It is beyond this study to assess to what extent the microphysical perturbations contribute to a better probabilistic forecasting skill compared to observation. Given the increasing importance of satellite observations used in convective-scale data assimilation the systematic impact of microphysical uncertainties will attract interest in future. Microphysical uncertainties strongly influence forecasts of cloud coverage and droplet sizes, both representing important ingredients used in satellite forward operators to compute synthetic reflectances (e.g. Scheck et al., 2020) to be used in data assimilation algorithms.

Chapter 4

Scale-dependent predictability and impact

The following Chapter is adapted from the publications titled "The impact of microphysical uncertainty conditional on initial and boundary condition uncertainty under varying synoptic control"(Matsunobu et al., 2022) and "Flow- and scale-dependent spatial predictability of convective precipitation combining different model uncertainty representations" (Matsunobu et al., 2024).

4.1 Background

Upon the systematic impact on spatially and temporally averaged variables in the previous chapter, the focus is extended to evaluating whether the impact helps to improve ensemble forecasts at scales more useful for our daily lives. This chapter narrows the target down to the impact on hourly precipitation forecasts in a scale-dependent manner, serving to answer the second research question **At what spatial scales do uncertainties impact?** The Fraction Skill Scores (Roberts and Lean, 2008) is employed to quantify the scale-dependent predictability (Sect. 2.4).

The investigation consists of two steps. First, in Sect. 4.2, the individual and synergistic impacts of IBC, MPP and PSP are assessed in a couple of case studies. This is followed by a systematic analysis of the impact of PSP in a long-term forecast dataset covering three months in the summer of 2021, presented in Sect. 4.3. The results are summarised in Sect. 4.4.

4.2 Quantification of the individual/synergistic impact

4.2.1 Case Study dataset v1: August 2020

To perform in-detail investigation of the scale-dependent impact of MPP, we the Case Study dataset v1 is used again for the results in Subsect. 4.2.4. Apart from that subsection all

results are derived using the Case Study dataset v2 described in the following subsections.

4.2.2 Case Study dataset v2: June 2021

Ensemble design

Table 4.1: Overview of ensemble experiments performed to gauge the relative and synergestic impact of different sources of uncertainty: acronym, ensemble size and perturbations.

Name	Ensemble size	IBC	PSP	MPP
IPM	120	✓	✓	✓
IM	120	✓		✓
IP	20	✓	✓	
I	20	✓		
P	20		✓	
M	6			✓

To gauge the individual and synergistic impact of different formulations of model uncertainty, the PSP scheme and parameter perturbations in the microphysics scheme, both in the presence of operational IBC uncertainties, we designed a 'Grand Ensemble' containing IBC uncertainty and two different flavours of model uncertainty. The 'Grand Ensemble' enables the inspection of the combined impact, but facilitates also an estimation of the individual impact of the PSP scheme and MPP by applying sub-sampling (as in Craig et al., 2022).

Here the 'Grand Ensemble' is a 120-member ensemble consisting of 20 different IBCs, the PSP scheme turned on and six realisations of microphysical uncertainty (denoted IPM, see Table 4.1). The IM ensemble lacks stochastic perturbations, while IP combines IBC uncertainty and the PSP scheme where each of the 20 ensemble members has different IBCs and different random seeds in the stochastic scheme as described in Chapter 2. Additionally we performed three ensembles containing the different sources of uncertainty individually. In the I ensemble IBC perturbations are the only source of uncertainty, the P ensemble has one IBC realisation (ensemble member 1 of I) but 20 different random seeds in the PSP scheme whereas the M ensemble solely includes six MPPs. The I ensemble can be seen as the reference ensemble in this part of the work. In order to assess the impact of microphysical uncertainty we used the two-moment bulk microphysics scheme (Seifert and Beheng, 2006) and turned off random parameter perturbations to purely focus on the impact given by the explicitly selected perturbations. Apart of these two differences the model set-up is identical to the trial dataset used in the first part. Since such an approach is a computationally expensive we restrict the numerical simulation to two cases representing varying convective equilibrium regimes taken from the trial period.

Weather situation and classification of cases

We chose two case studies from the period of the trial run in summer 2021 representing both convective forcing regimes. Among the days characterised by weak convective forcing, 10 June was one of the most typical non-equilibrium days that showed a large impact of the PSP scheme on precipitation. On that day, the atmospheric flow over Germany was characterised by weak north-westerly winds due to a small geopotential gradient. Scattered convection was triggered around noon and reached its maximum intensity around 1400 UTC. The daily accumulated rainfall exhibits a popcorn-like pattern typical for weak convective forcing (Fig. 4.1a,b). In the evening the atmosphere was stable again, ending the well-defined diurnal cycle of convection.

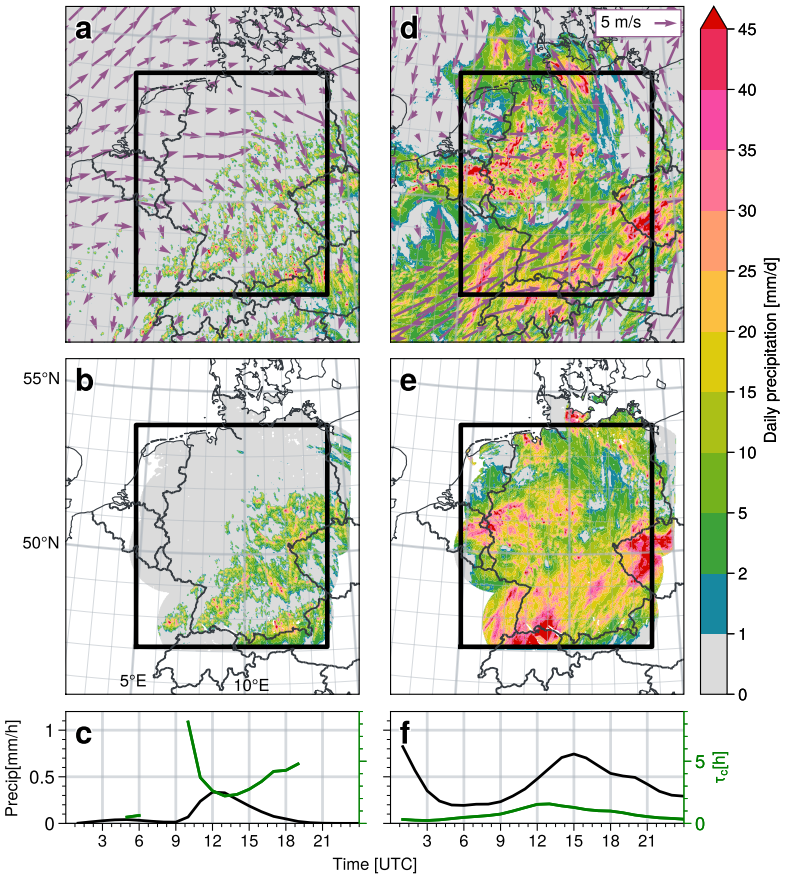

Figure 4.1: Characteristic fingerprint of precipitation during different convective forcing regimes in central Europe illustrated by daily accumulation of precipitation for a weak forcing case on 10 June (a, b) and a strong forcing case on 29 June (d,e) of an ensemble member of the TR ensemble (a, d) and radar observations (b, e). Vectors depict the wind velocity at 850 hPa pressure level. The black rectangle indicates the verification domain. The bottom row (c,f) shows hourly time series of ensemble mean precipitation (black) and the convective adjustment time scale τ_c (green).

In contrast, weather on 29 June represents a typical day of the strong convective forcing regime, for a number of reasons. Although 29 June is not part of the equilibrium composite of the trial period, it was one of the most representative meteorological situations of its kind, with a strong geopotential gradient, strong south-westerly flow (Fig. 4.1c) and the largest accumulated precipitation in the whole summer period, according to radar observations (Fig. 4.9). The development of a mesoscale convective system along the cold front in southern Germany caused an outbreak of high impact weather with severe winds, hail and heavy precipitation. A mean daily area-averaged value of τ_c amounting to less than two hours clearly classifies this day as being in equilibrium (Fig. 4.1f and 4.9). This case was also part of an intensive observation period (IOP 5, 28-30 June) of the Swabian MOSES field campaign (Kunz et al., 2022).

4.2.3 Area-averaged precipitation amount and spread

First, we inspect the impact of IBC, PSP and MPP on ensemble mean precipitation and ensemble spread. Figure 4.2 shows ensemble- and area-averaged hourly precipitation amounts and its area-averaged spread for four experiments: the pure IBC perturbed ensemble (I), the combined IBC and PSP ensemble (IP), the IBC and MPP perturbed ensemble (IM) and the ensemble containing all uncertainty representations (IPM). On the non-equilibrium day, the onset of precipitation is earlier and the peak of precipitation intensity is enhanced when adding PSP compared to pure IBC uncertainty (Fig. 4.2a). This is attributable to the more effective trigger mechanisms introduced by the PSP scheme. The addition of PSP enhances also the spread. Combining IBC and MPP (IM ensemble) gives the larger and prolonged spread. The spread is largest for the IPM ensemble during convection, followed by the IP and IM ensemble, respectively. The I ensemble exhibits the lowest spread and the peak is reached about one hour later than in ensembles IP and IPM. The effect of both model uncertainties, PSP and MPP, complement one another in terms of ensemble spread. During strong convective forcing the area-averaged precipitation amount is governed by IBC and only marginally influenced by any kind of model uncertainty representation. Such a dominant role of IBC on precipitation forecasts was previously shown, for instance, by Johnson and Wang (2020). Spread is hardly changed by adding PSP to IBC, and is even slightly reduced in the late afternoon (Fig. 4.2b), while adding MPP enhances the spread considerably. Similar to the non-equilibrium case, the change in spread given by adding PSP and MPP shows a qualitatively additive character in IPM. Both cases illustrate that the synergistic impact of PSP and MPP (IPM ensemble) can be beneficial since both model uncertainty representations partly compensate the respective deficiencies.

To assess whether the earlier onset of convection due to PSP and the prolonged spread due to MPP are caused by their individual impact rather than the interaction with IBC perturbations, two additional ensembles are examined: the P and M ensembles. In the P ensemble, the only source of uncertainty is the PSP scheme, whereas the M ensemble contains only MPP. Both ensembles are run with one set of IBC (member 1 of the I ensemble), which does not allow for a quantitative comparison in Fig. 4.2. However, the

reader can qualitatively compare these ensembles in Fig. 4.7. The P ensemble shows an earlier intensification of precipitation and spread (not shown), which is in line with Leoncini et al. (2010), who found that large perturbations in the planetary boundary layer lead to an earlier growth of perturbations in precipitation. The M ensemble shows a smaller but slower decay of spread at later stages of convection, as shown by Barthlott et al. (2022a). This indicates that PSP and MPP act independently at different stages during the diurnal cycle of convection.

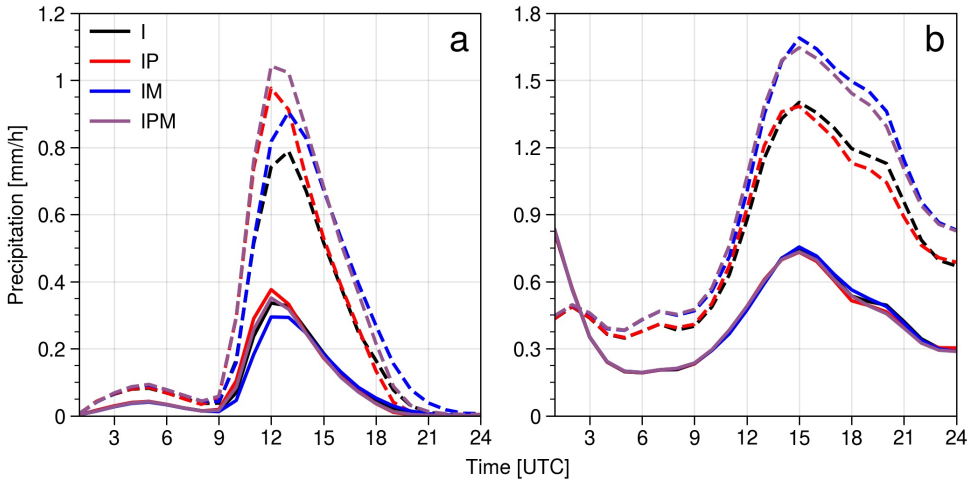

Figure 4.2: Ensemble- and area-averaged hourly precipitation (solid) and its area-averaged spread (dashed) over Germany for a weak forcing case on 10 June (a) and for a strong forcing case 29 June (b) representing varying convective forcing regimes.

4.2.4 Spatial variability given by MPP

To address the question of how IBC and microphysical uncertainties affect convective precipitation on different spatiotemporal scales we now move from area averages to the kilometre scale and from daily to hourly accumulations. The fractions skill score (FSS; Roberts and Lean, 2008) and its variant believable scale (Dey et al., 2014; Bachmann et al., 2020) are used to objectively assess differences in spatial variability caused by different sources of uncertainty. But first we apply subjective visual inspection on selected precipitation fields to illustrate differences.

In Fig. 4.3 a snapshot of hourly precipitation over central western Germany (blue box in Fig. 3.2a) for the weak forcing case (11 August) at 16 UTC exemplifies the different impact of IBC and microphysical perturbations. This day is chosen because of the stronger impact of the perturbations during weak synoptic control, and 16 UTC represents the time of maximum afternoon precipitation within the diurnal cycle of convective precipitation (see Fig. 3.2c), and the displayed subdomain clearly depicts the typical popcorn-type precipitation structure. In Fig. 4.3 the transient character of individual cells is juxtaposed for four different experiments: three of them share the identical IBC (panels a, b and c),

CCN concentration (panels a, b and d) and shape parameter of CDSD (panels a, c and d), respectively.

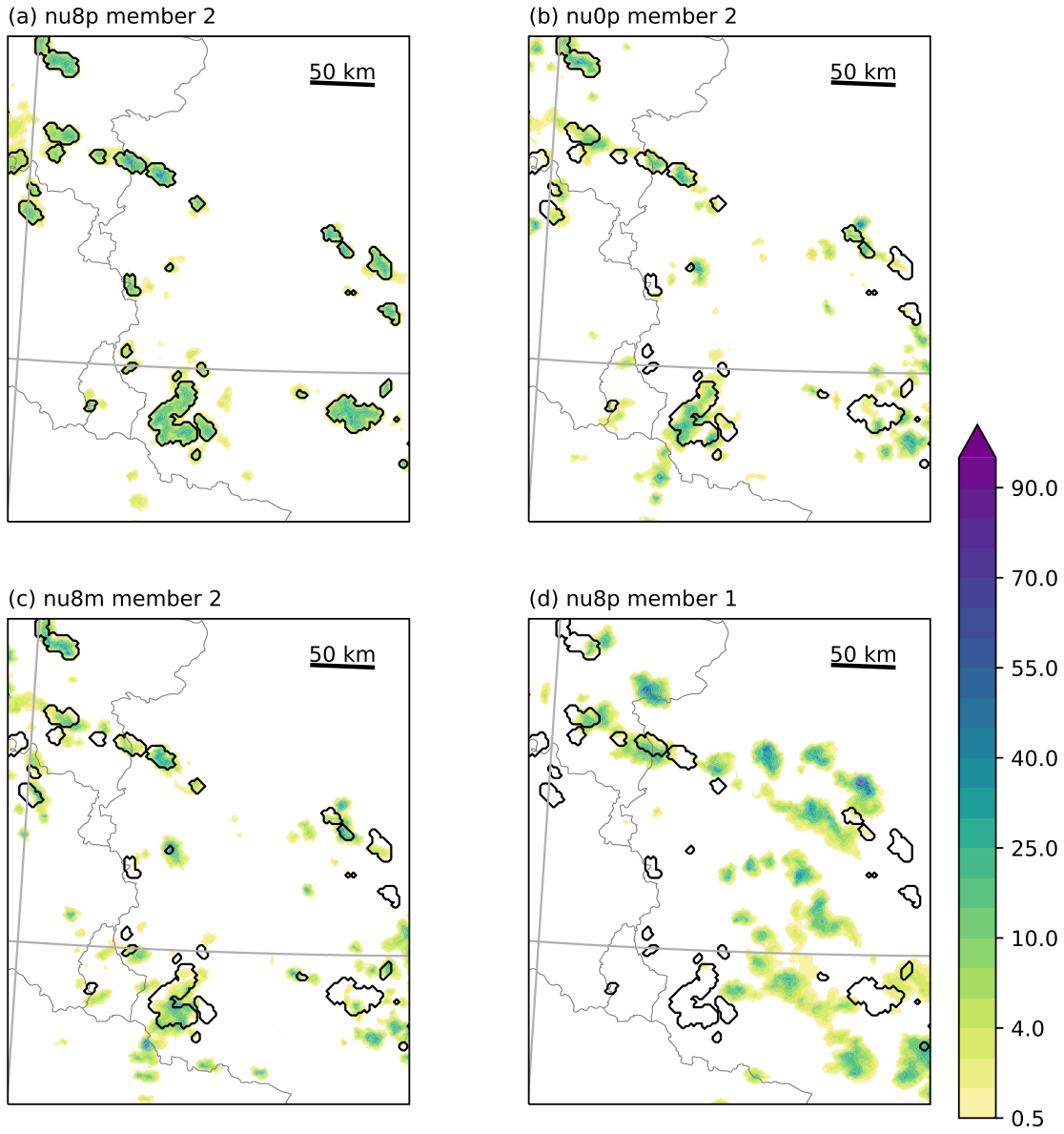

Figure 4.3: Snapshot of hourly precipitation at 16 UTC for the weak forcing case (11 August). Member 2 of IBC sub-ensembles (a) *nu8p*, (b) *nu0p*, (c) *nu8m* and (d) member 1 of *nu8p* in the central western part of Germany (see blue box in Fig. 3.2). Black contours indicate grid points that have a larger value than the 99th percentile value in the *nu8p* sub-ensemble of member 2. The red circle in (a) indicates single convective cells discussed in the text.

At first glance, it becomes evident that the microphysical perturbations result in a

similar rainfall distribution (Fig. 4.3a, b, c), whereas the member driven with different IBC shows a considerably different rainfall field (Fig. 4.3d). The direct comparison of the location of intense precipitation caused by the different perturbations relative to the 99th percentile of simulation *nu8p* (black contours in Fig. 4.3) shows that convective cells of simulations *nu0p* (broad CDSD, polluted) and *nu8m* (narrow CDSD, maritime) are either at the same location or in close vicinity. Some weak rain cells (e.g. southeast of Luxemburg, red circle in Fig. 4.3a) are intensified by decreasing CCN and shape parameters of CDSD, thus in agreement with the spatiotemporal integrated rainfall signal discussed in the previous section. Positions of strong rain cells are shifted by the CCN perturbation at a scale of 20-30 kilometres, whereas an increase in the shape parameter of CDSD hardly shows a clear impact. The visual inspection of many scenes of hourly rainfall caused by convective cells confirms the systematic behaviour of microphysical perturbations with stronger precipitation with low CCN concentration and broad CDSD shapes (not shown).

To briefly summarise the visual inspection, we can state that in polluted CCN conditions both CCN and CDSD perturbations impact the spatial variability at almost the same scale. While microphysical perturbations keep the general spatial structure, IBC perturbations largely alter the position of convective cells. Thus microphysical perturbations primarily impact the precipitation amount by changing their precipitation intensity rather than by feedback on dynamical fields and triggering new cells. Visual inspection of rainfall patterns of the strong forcing case results in similar findings: minor shifts of rain cells in microphysics sub-ensembles and a smaller impact of CDSD perturbations (not shown).

To quantify the spatial (dis-)agreement of hourly precipitation fields in the various simulations we employ the FSS, a spatial score that shows the similarity between two binary fields (denoted A and B, two distinct sub-ensemble members in our case), within a predefined neighbourhood scale. The definition of the FSS is given by

$$FSS = 1 - \frac{\sum(f_A - f_B)^2}{\sum f_A^2 + \sum f_B^2} \tag{4.1}$$

where f_A and f_B represent the fraction of rainy grid points in fields A and B, respectively, at which the precipitation amount is above a certain threshold value. The second term on the right-hand side is the ratio of the mean squared error (MSE) of the fraction fields A and B to the maximum possible MSE (Roberts and Lean, 2008). If the number of grid points with a value of 1 within a certain neighbourhood of a grid point is equal between two fields, the FSS is 1.0, which means the compared two fields are identical. FSS becomes smaller as the difference between two fields gets larger, and it becomes 0.0 when only one of the fields has values and the other has a complete miss in the respective neighbourhood. In this study, we use the 99th percentile of hourly precipitation as the threshold to generate a binary field to take into account the strong diurnal cycle of rainfall intensity and to keep the number of grid points used for FSS calculation constant. The 99th percentile is useful to capture the position of convective cells (see contours in Fig. 4.3). The FSS is calculated over Germany with neighbourhood sizes varying from 2.2 km (1 grid point) to 563.2 km (256 grid points). Since FSS is a score calculated between two fields, we need to carefully consider how to compute an ensemble FSS. Following Dey et al. (2014), we

calculate the FSS for all combinations of ensemble members belonging to a sub-ensemble. For instance, FSSs for an IBC sub-ensemble (with 20 different IBC) can be calculated 20 * 19 / 2 = 190 times. Since there are 9 IBC sub-ensembles in this study, the number of overall FSSs that shows the impact of IBC perturbations is 190 * 9 = 1710. Accordingly, the numbers of FSSs for combined microphysics, CCN and CDSD sub-ensembles are 720, 180 and 180, respectively. Mean values of the FSSs are shown in Figs. 4.4 and 4.5 to objectively represent the spatial variability given by various kinds of uncertainties.

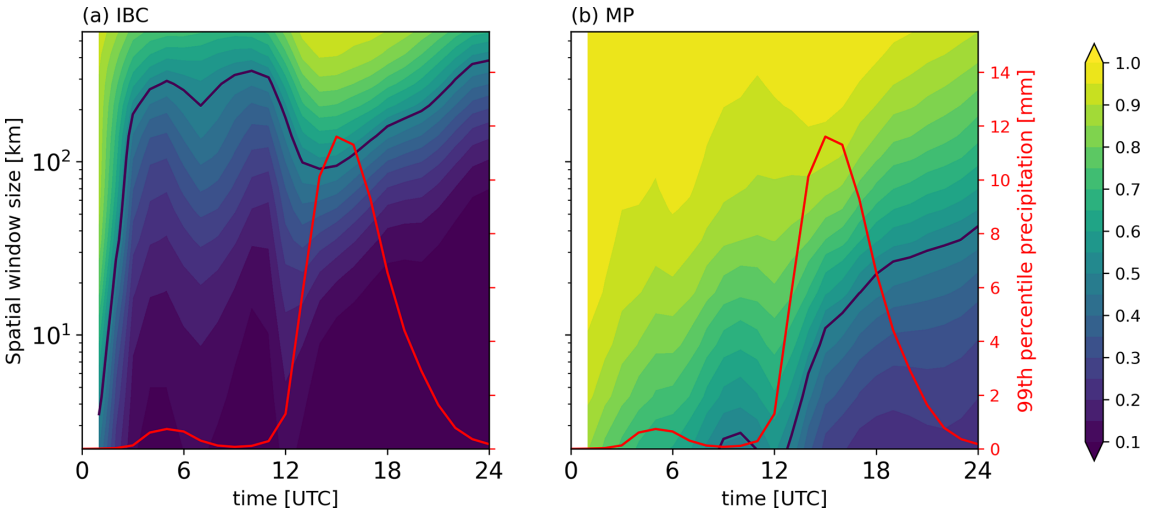

Figure 4.4: Ensemble mean FSS values of hourly precipitation calculated across scales ranging from 2 to 560 km in the German domain for the weak forcing case 11 August. The IBC sub-ensembles' mean FSS is depicted in panel (a) and the combined microphysics sub-ensembles' mean FSS in panel (b). The black lines show believable scales of mean FSS. The red lines (right axis) show the time series of mean 99th percentile value of hourly precipitation.

In addition, we use the believable scale (Dey et al., 2014; Bachmann et al., 2020) to characterise a typical length scale that estimates the spatial difference between two fields. The believable scale is defined as the neighbourhood size when the FSS exceeds a threshold defined by $FSS \geq 0.5 + \frac{f_0}{2}$, where f_0 is the fraction of grid points considered in the FSS calculation (the 99th percentile threshold gives $f_0 = 0.01$). Since the FSS is applied on precipitation fields above the 99th percentile values, the believable scale can be considered in this study as a scale showing how large a mismatch of intense convective cells is.

Time-space diagrams of the ensemble mean FSSs given by IBC and combined microphysical uncertainty are depicted in Fig. 4.4 for the weak forcing case. Low FSS values represent large spatial deviations between the location of intense convection, hence a larger spatial variability. The variability due to the IBC perturbations is considerably larger than the one forced by combined microphysical perturbations. However, and typical for days with weak control, convective precipitation only forms in the late morning (see e.g. time series in Fig. 3.2c and red line depicting the 99th percentile of hourly precipitation in

Fig. 4.4). The value of the 99th percentile of hourly precipitation amounts to 1 mm/h at 12 UTC and precipitation is mostly negligible before. Interestingly, at the onset of convective precipitation at 12 UTC the believable scale exhibits a dip and the spatial variability decreases to slightly less than 100 km and thereafter continuously increases throughout the convective period until the evening. The reduction of the spatial variability in the afternoon, representing co-locations of convective cells, is constrained by steady, non-perturbed factors forcing the dynamical fields involved in cloud and precipitation formation like orography. After 22 UTC the hourly precipitation rates amount again to less than 1 mm/h and the corresponding believable scale exceeds 200 km like before the onset of convection at night and in the morning. In contrast, the spatial disagreement caused by combined microphysical perturbations is smaller and the mean believable scale amounts to only 16 km at the peak of precipitation at 16 UTC (Fig. 4.4b). Apparently, the impact of microphysical perturbations on precipitation acting on many pathways needs time and starts at a much lower spatial scale than IBC perturbations.

At first sight, individual perturbations of CCN and CDSD show a similar growth of FSS as the combined microphysical perturbations (Fig. 4.4b and Fig. 4.5). Close inspection reveals that the believable scale of precipitation caused by CCN perturbations (black line in Fig. 4.5b) starts to increase at the onset of the precipitation (at 12 UTC), one hour before that of the CDSD perturbations (Fig. 4.5a). The CDSD believable scale grows more slowly and is always smaller (roughly 50%) than that of combined microphysical perturbations. Since changes in CCN have a direct influence on the cloud condensation process, while the shape parameter of CDSD affects ensuing microphysical processes, this time shift is plausible. Interestingly, the CCN perturbed believable scale reaches 40 km after 22 hours, the same length scale as the believable scale of the combined microphysical perturbations. In contrast to the impact on precipitation amount, combining two sources of microphysical uncertainty does not increase the spatial variability.

The uncertainty of CCN concentrations has a larger impact than the shape parameter of CDSD on the spatial variability of intense precipitation cells. Now we can ask if this behaviour is by chance and if this finding holds for other thresholds or percentiles, respectively. For this reason, we performed additional white noise (WNoise) ensemble simulations with 20 different IBC but only for the 'default' microphysics configuration ($nu0c$) to examine whether the spatial variability caused, for instance, by microphysical perturbations differs from the impact of random, tiny differences in the temperature field. Following the method of Selz and Craig (2015) the virtual potential temperature field is perturbed by a non-biased Gaussian noise with a standard deviation of 0.01 K at all grid points of the entire model atmosphere at an initial time. The comparison of the microphysically perturbed ensemble with a pure white noise (WNoise) experiment shows a similar onset and increase of spatial variability (Fig. 4.5c). The spatial variability caused by CCN and CDSD perturbations is, however, larger than the effect of the WNoise perturbations. At 16 UTC, the mean FSS of WNoise simulations is close to 1 at scales larger than 80 km, and the believable scale is about 5 km. Thus the effect of microphysical uncertainty on the spatial precipitation fields is systematically exceeding the effect of tiny errors at the initial time in the WNoise experiment. Less intense precipitation cells detected by the

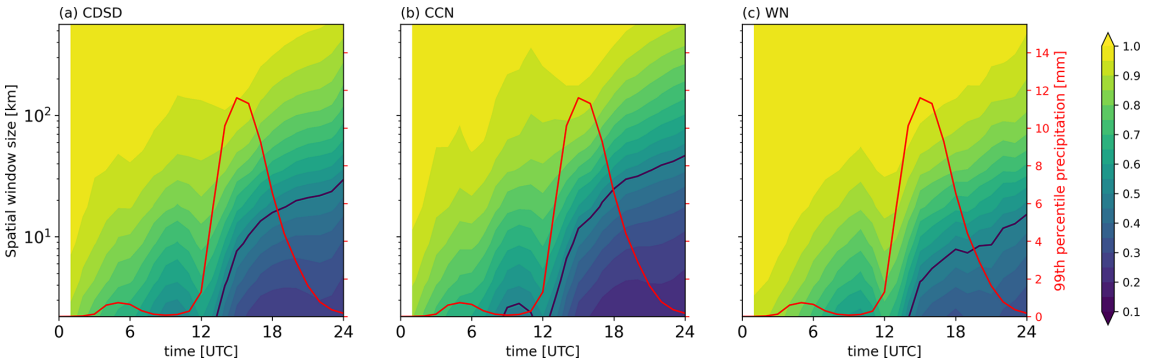

Figure 4.5: As Fig. 4.4, but for the (a) CDSD, (b) CCN and (c) WNoise sub-ensembles.

95th percentile threshold indicate a similar albeit slightly smaller variability due to IBC and microphysical perturbations (not shown). Using a 90th percentile threshold on hourly precipitation results in values lower than 0.1 mm at all forecast hours and gives no extra information.

To further elucidate the combined microphysical perturbations and the interdependence of one perturbation (say CCN) when the other (CDSD) is kept constant in the presence of IBC uncertainty, time series of all believable scales calculated between every combination of ensemble members are illustrated in Fig. 4.6. The bold lines in Fig. 4.6a clearly reveal that CDSD perturbations result in spatial variability at different length scales depending on a certain fixed CCN concentration during weak synoptic control. In clean air conditions (maritime aerosol content, dark blue lines in Fig. 4.6a), the mean believable scale attains 10 km roughly 3 hours after the onset of the believable scale's growth. At 22 UTC, towards the end of the diurnal cycle, the value increases to 15 km. On the other hand, for polluted conditions (dark red and green lines), the mean believable scales attain larger values, 15 km at 16 UTC and 30 to 40 km at 22 UTC. The mean length scale of disagreement given by the CDSD perturbations in polluted conditions (high CCN concentrations) is twice as large as in clean conditions (low CCN concentrations). Note, however, that there is big variability among pairs of ensemble members, hence the IBC dependence is larger than the impact of the background CCN condition. A similar systematic dependence can be found for the CCN perturbations' impact with different fixed CDSD shape parameters. The mean believable scale with the broadest CDSD (lightest grey lines in Fig. 4.6b) reaches 10 km at 16 UTC and 50 km after 22 hours of lead time. With the narrowest CDSD (black lines), the mean believable scale of CCN perturbations is 20 km at 16 UTC and increases to 100 km later. Interestingly, the mean believable scale with the narrowest CDSD is by a factor of 2 larger than the broadest CDSD. This relationship is similar to that found in spatially averaged precipitation amounts, namely polluted CCN and narrow CDSD conditions lead to larger variability (Fig. 3.4).

In strong synoptic control, the situation is slightly different (Fig. 4.6c,d). The believable scales only start to grow from 7 UTC onwards, and the mean values finally reach a neighbourhood size of 30 km at 22 hours lead time. This monotonic pattern of the per-

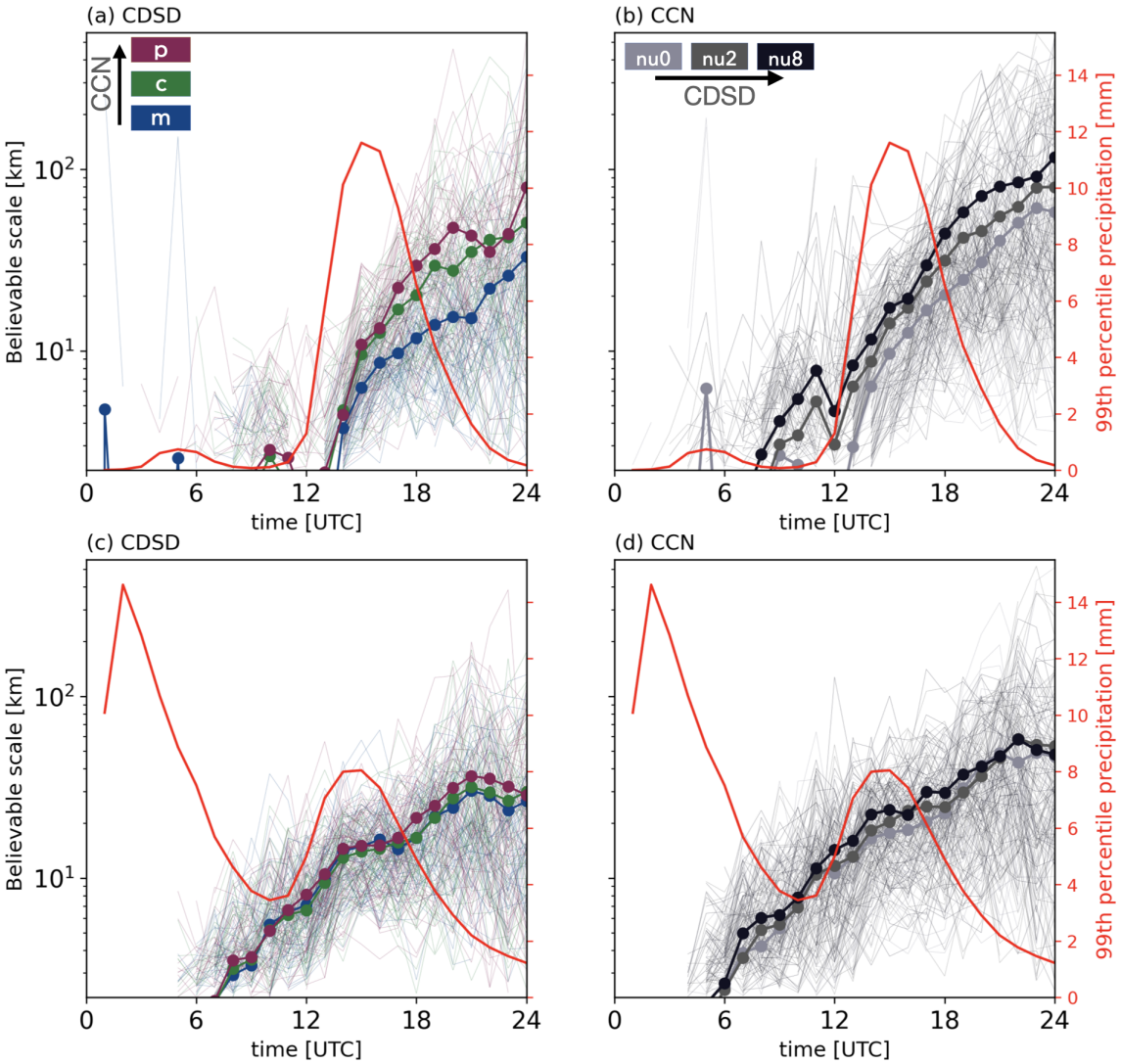

Figure 4.6: Time series of FSS believable scales of hourly precipitation for every combination of (a) the CDSD and (b) CCN sub-ensemble for the weak forcing case in the German domain. In (a) blue, green and red lines indicate simulations with maritime, continental and polluted CCN content, respectively. In (b) light grey, dark grey and black lines indicate scales with the broad, intermediate and narrow CDSD. Bold lines with circles indicate mean values of FSS believable scales sharing the same perturbation. The red lines (right axis) show time series of mean 99th percentile value of hourly precipitation. Panels (c) and (d) show the results for the strong synoptic-forcing case.

turbation growth is similar to the weak forcing case. However, the mean believable scale for clean CCN conditions is larger than for the weak forcing case at 22 UTC (dark blue bold lines in Fig. 4.6a and c). There is no systematic difference in the mean believable scale caused by CDSD perturbations in the presence of various, yet fixed CCN concentrations (Fig. 4.6c). On the other hand, given narrower CDSD, the CCN perturbations cause a slightly larger spatial variability (Fig. 4.6d). Nevertheless, the difference between the broadest and narrowest CDSD simulations is less pronounced in comparison to the weak forcing case (10-15 km difference in strong control versus 30 km in weak control at 22 UTC). It is interesting to note that the impact of the microphysical perturbations on the spatial precipitation pattern only starts to appear in FSS after 7 hours of lead time, although there is continuous rainfall since forecast initialization during the strong forcing case. In the first hours of the simulation spin-up effects and the adjustment to the driving coarser-scale model are still at work, which dampens the impact of the microphysical uncertainties (see, e.g., Barthlott et al., 2022b). Thus, microphysical perturbations need a longer spin-up time than IBC perturbations to modulate dynamical fields eventually resulting in precipitation at different locations (see Fig. 4.6c,d).

Note that there is a difference between the believable scale of a 'mean FSS' (e.g. black line in Fig. 4.4) that represents a scale of (dis-)agreement given, say, an ensemble mean FSS value and the mean over many believable scale values of paired member-to-member comparisons (Fig. 4.6). The ensemble mean FSS is useful for an intercomparison of the average impact given by different perturbations in general, whereas the mean of member-to-member believable scales (Fig. 4.6) provide a scale of actual (dis-)agreement of certain scenes, for example, the precipitation patterns shown in Fig. 4.3.

4.2.5 Spatial variability given by PSP

Building upon the beneficial impact of the PSP scheme shown in the systematic assessment, we strive to inspect the influence of additional sources of model uncertainty inserted into the full CPEPS. Due to computational constraints, we limit the analysis on selected cases representative of different flow situations.

To set the scene for the scale-dependent examination of the relative influence of three uncertainty representations in the 'Grand Ensemble' (IPM, see Table 4.1) we first display the time series of spatial spread due to the individual uncertainties (I, P and M ensembles) for both cases in Fig. 4.7. The weakly forced case exhibits a typical diurnal cycle with the most intense rainfall occurring soon after noon. The characteristic upscale error growth caused by displacements of convective cells is quantified by the spatial spread. Its temporal evolution confirms this steady increase (coloured tiles in Fig. 4.7a). In the equilibrium regime, IBC clearly represents the dominant source of uncertainty in terms of amount (solid lines in Fig. 4.2b) and in terms of location of heaviest precipitation (Fig. 4.7d). The spatial spread of precipitation grows upscale shortly after model initialisation fuelled by strong nighttime rainfall and exhibits large displacements throughout the day.

The spatial spread of model uncertainty represented by pure PSP is displayed in Fig. 4.7b,e. During weak convective forcing the impact of PSP is in general stronger

and conditioned by the daily cycle of convection. The perturbations added in the boundary layer start having an impact on precipitation at the onset of convection, when thermals start to trigger convective cells. As soon as that happens, the spatial scale of spread quickly grows upscale and reaches the maximum extent in the evening when forecast precipitation ceases. The magnitude of displacements in the P ensemble at the peak time of precipitation (12–14 UTC) is in parallel with the impact of the SBL scheme found by Flack et al. (2021). In strong convective forcing the effect of PSP is comparatively small but grows continuously at a slower rate driven by continuing rainfall.

The M ensemble shows different growth patterns depending on varying convective forcing regimes. During weak forcing of convection, spatial spread grows from the smallest scale towards larger scales as for P, but the amplitude is smaller (Fig. 4.7c). During strong forcing MPP impact the spatial spread at small scales from the early morning onward but this impact is largely comparable with that of the PSP scheme thereafter (Fig. 4.7f). However, after 19 UTC the MPP impact spreads across scales in contrast to the PSP impact that stays below 80 km. Compared to the impact of PSP, MPP induce similar upscale growth in the weak convective forcing case, but at a slightly later stage, after the onset of convection, since MPP only start acting when convection is ongoing, as shown in Matsunobu et al. (2022).

4.2.6 Synergistic impact on spatial error and spread

Finally, the impact of combining PSP and MPP with IBC on the spatial error and spread can be discussed by analysing differences of the IP, IM, and IPM ensembles from the I ensemble (Fig. 4.8). Results are displayed for the weak convective forcing regime. To begin with, the time series of the 95^{th} percentile of precipitation for the ensembles show again that both model uncertainties compensate each other in the afternoon (compare blue bold lines in Fig. 4.8a,c,e), similar to the time series of spread depicted in Fig. 4.2a. Turning towards the spatial pattern of hourly precipitation, we find that the contribution of the PSP scheme (Fig. 4.8a, b) resembles the one discovered in the composite analysis (Fig. 4.11a, b). The diurnal cycle of precipitation is shifted earlier by the PSP scheme, leading to a significant error reduction at the onset of convection (1000 UTC) at all scales. However, there is a slight spatial error increase in the afternoon. A similar signal can also be found in Fig. 4.11a discernible as slightly lighter colors at 1600–1700 UTC, presumably due to the earlier decay of convection. However, on average the PSP scheme systematically improves the spatial skill as shown in Fig. 4.11a. The PSP impact on spatial spread, on the other hand, is systematically increasing and growing upscale from about 10 to 80 km after the onset of precipitation, in agreement with earlier findings (see Fig. 4.10). The impact of adding MPP on spatial error is opposite from that of PSP. Including MPP reduces precipitation intensities on average and also indicates a detrimental impact on spatial error that is largest at the early stages of convection (1000–1200 UTC) rather uniform across all scales (Fig. 4.8c). Subsequently, the change in spatial error is still slightly positive on scales larger than 80 km and neutral on scales below 50 km. On the other hand, adding MPP increases the spatial ensemble variability almost throughout the forecast at all scales

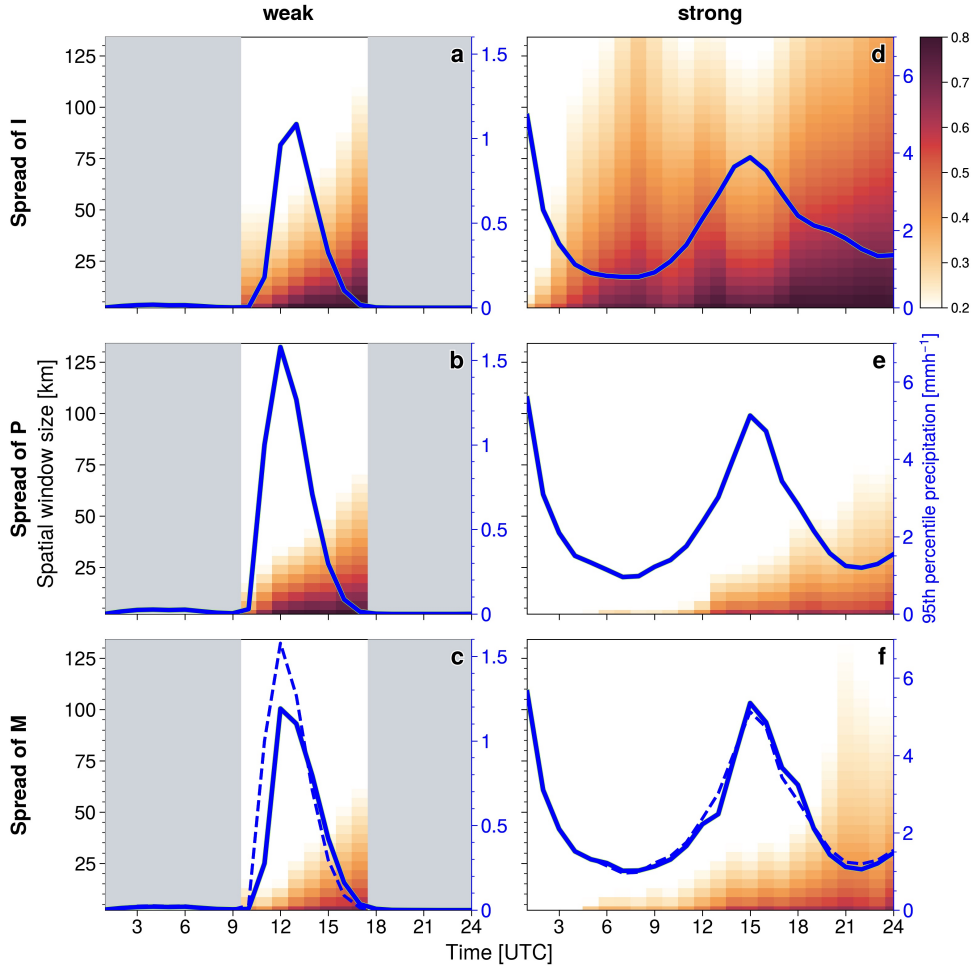

Figure 4.7: Spatial spread in ensemble I (top), P (middle) and M (bottom) on 10 June (left) and 29 June (right). The blue lines indicate the 95^{th} percentile values of hourly precipitation of I (a,d), P (b,e continuous; c,f dashed) and M (c,f continuous) ensemble. Masking is applied when the observed or forecast precipitation amount is below 0.01 mm/h.

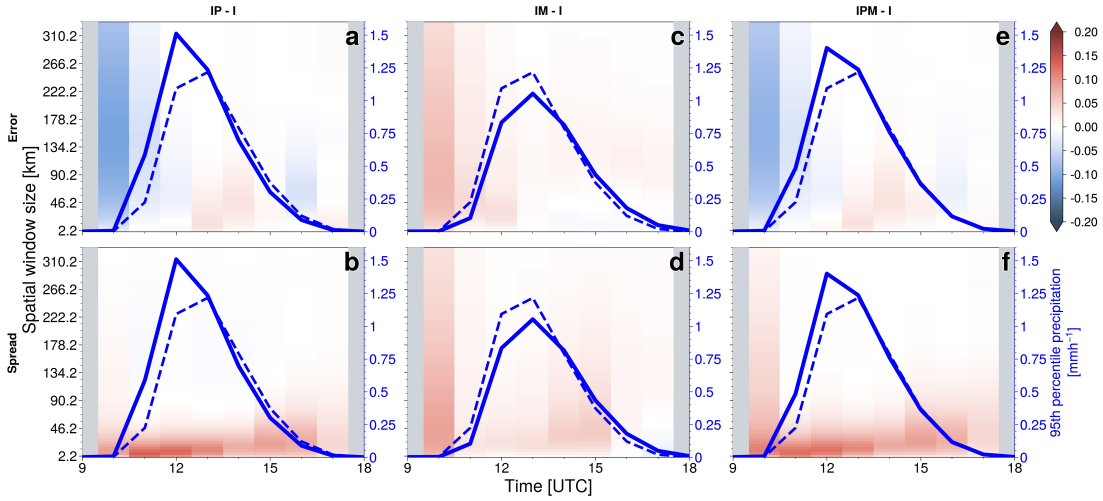

Figure 4.8: Time series of difference in spatial error (top) and spread (bottom) between combined ensembles and the I ensemble for 10 June. Panels show the change given by adding PSP (a,b), by adding MPP (c,d) and by adding both PSP and MPP (e,f). Blue (red) shading indicates a reduction (increase) in spatial error and spatial spread, respectively. The blue lines indicate the 95^{th} percentile values of hourly precipitation of the reference I ensemble (dashed) and IP (a,b), IM (c,d) and IPM (e,f) ensembles (bold).

(Fig. 4.8d). In this case, the up-amplitude effects of MPP are relatively independently of scales, potentially improving the location of the strongest precipitation (the 95^{th} percentile) on scales below 50 km but do not propagate towards larger scales as seen with PSP.

We hypothesise that this scale-independent effect of MPP is caused by the design of the MPP being fixed in space and time (i.e. fully correlated) and not being scaled in direct response to the occurrence and scales of physical processes. This is beneficial at smallest scales where spatial error is large and the ensemble is overconfident, but at the same time can deteriorate forecast skill at larger scales where IBC perturbations dominate and the spread to skill ratio is already converging. Nevertheless, the present implementation of MPP improves the overall spatial spread to skill ratio as it generally increases spatial variability, which can add value in an overconfident ensemble. A refined configuration of MPP, or a combination with other perturbations that better represents spatio-temporally varying microphysical uncertainty ought to be pursued in future.

When comparing the contributions of PSP to those of IBC, we find that the gained spatial variability remains below 50 km, at a smaller scale than the variability obtained with IBC. This is similar to the effect of SBL in Clark et al. (2021) and Flack et al. (2021). However, PSP changes the time evolution of precipitation and largely reduces the spatial error at the time of the onset of convection. These improvements are key advantages of the PSP scheme and cannot be achieved using current postprocessing methods. On the other hand, MPP have a broad impact on the spatial spread throughout time and scales. They are systematically shifting the model climate rather than sampling the impact of random

errors, as discussed in McTaggart-Cowan et al. (2022). Although this causes an increase in spatial variability, other metrics should be used to assess whether such an approach improves other aspects of the forecast, such as the bias in amplitude.

Overall, the joint impact of PSP and MPP in terms of spatial error and spread appears to be additive (Fig. 4.8e, f) with PSP being the primary source of uncertainty. Precipitation is triggered earlier and its intensity is higher. The earlier decay of precipitation is attenuated by MPP. Spatial error is reduced at convection initiation, while spatial spread is increased across time and scales. Hence the spatial error-spread relationship is significantly improved compared to the pure IBC ensemble. These results imply that the impact of both sources of model uncertainty (that is PSP and MPP) on spatial variability of precipitation are non-correlated and rather orthogonal suggesting that these perturbations can effectively work together. However, alternative adaptive and statistical postprocessing methods could be used instead to represent variability induced by the PSP scheme, and might also remedy the undesired impact of MPP at scales larger than 100 km, as shown in Blake et al. (2018) and Flack et al. (2021).

4.3 Systematic analysis of the flow-dependent predictability and impact

4.3.1 The trial run in DWD

Ensemble design

The long-term assessment of the spatial predictability builds on the ensemble forecast dataset of the trial run (Puh et al., 2023). This dataset facilitates a systematic evaluation of the flow- and scale dependence during the three-month period June to August 2021. The reference ensemble (denoted with trial reference, TR) is identical with the operational 20-member ICON-D2-EPS, driven by operational ICs provided by ICON-D2-KENDA and LBCs provided by ICON-EU-EPS, using the one-moment bulk-microphysics scheme and including the random physics parameter perturbations. In a second parallel ICON-D2 ensemble we use this ICON-D2 ensemble configuration and additionally turn on the PSP scheme (denoted TP henceforth). All simulations were initialised daily at 0000 UTC with 24 hour forecast lead time and were performed on DWD's High Performance Computer (more details in Puh et al. (2023)).

Weather situation and classification of cases

The different convective environments during the three-month period are objectively classified with the convective adjustment time scale (τ_c) diagnostic (Section 2.3). After excluding 12 days without precipitation there remain 80 days to be arranged in distinct convective forcing regimes. To be consistent with the original publication on the trial experiments (Puh et al., 2023), we use the τ_c diagnostic in a relative sense and group days with daily

mean τ_c values belonging to the upper (lower) 20 % of the distribution throughout the summer season into the weak (strong) convective forcing regime, respectively. This results in 16 days each falling into these two categories. The remaining 48 days in the middle of the spectrum of τ_c values make up the intermediate category, for which a primary type of convective forcing is not unequivocally detectable. Figure 4.9 depicts the classification of the days along with daily accumulated and area averaged ensemble mean precipitation amounts and radar-observed quantitative precipitation estimates. While the observed and forecast daily precipitation amounts reasonably agree, there are large variations of daily totals from day to day. Days governed by equilibrium conditions predominantly exhibit larger rainfall accumulations than non-equilibrium days. To enable a fair comparison of the spatial predictability we need to take these amplitude changes into account. That is accomplished by using a percentile threshold in the FSS calculation as outlined in Section 2.4a.

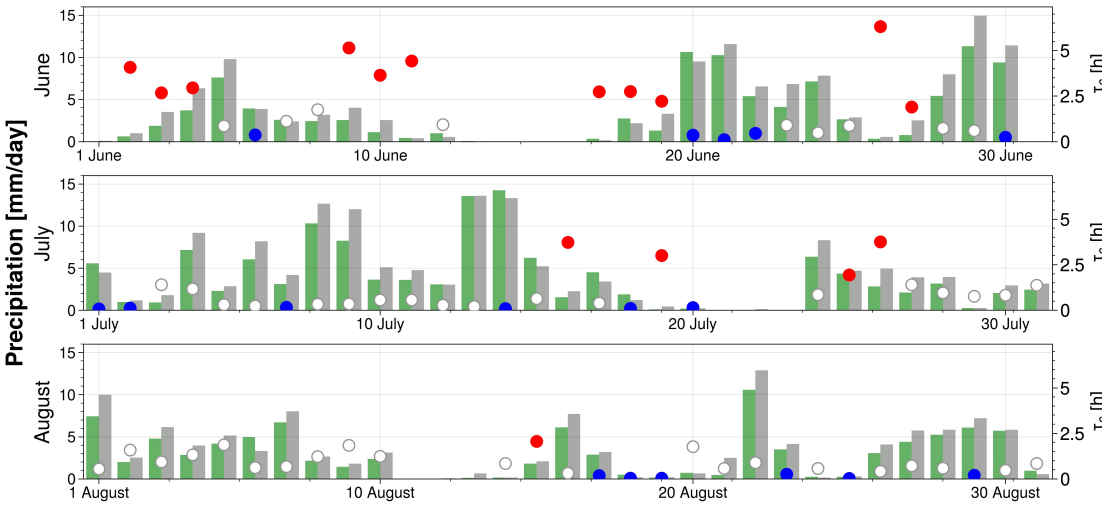

Figure 4.9: Time series of June, July and August 2021 illustrating the day-to-day variability of 24-h accumulated precipitation (bars) and convective adjustment timescale τ_c (dots). The colours of the dots represent weak (red), intermediate (white) and strong (blue) forcing regimes (see text for details). Green bars depict TR ensemble mean and grey bars the radar-observed daily area-averaged rainfall.

4.3.2 Flow-dependence of spatial predictability of hourly precipitation

Having objectively classified convective weather into different categories, we can inspect the flow-dependent spatial error and spread now. The operational ICON-D2-EPS forecasts (TR ensemble) show the typical diurnal cycle during weak convective forcing with the 95^{th} percentile value of hourly precipitation attaining highest intensities in the afternoon (0.3 mm/h, Fig. 4.10a). Not surprisingly the largest spatial errors occur on the smallest scales (high error values shown in dark colours in Fig. 4.10a). However, the spatial error

exhibits a clear diurnal cycle, having a maximum around 0900–1000 UTC at the time of triggering of moist convection (eDS approx. 170 km) followed by a minimum at peak afternoon precipitation (eDS approx. 70 km). Thereafter the spatial error increases again as the afternoon precipitation fades. At first sight the spatial spread behaves similarly as the spatial error (compare colours in Fig. 4.10a and b) but with lower values, that is smaller spatial variability. There is a relative maximum of the spatial spread at the time of convection initiation (0800–0900 UTC) followed by slight decrease in the afternoon during strongest convective activity. In particular the spread at small scales grows slower than the spatial error. In terms of the displacement scale, the sDS shows a modest temporal evolution and amounts to about 40 km, hence smaller than the eDS (the bold line is below the dashed line in Fig. 4.10b). A gap of several tens of kilometres between the sDS and eDS is evident throughout the forecast, even after the precipitation maximum in the afternoon, which indicates that the ensemble is spatially under-dispersive. Hence the ensemble forecast is spatially overconfident, that is the spatial spread is too low compared with the spatial error. This lack of spatial spread suggests that the perturbations in ICON-D2-EPS are sub-optimal and that additional sources of model uncertainty ought to be added.

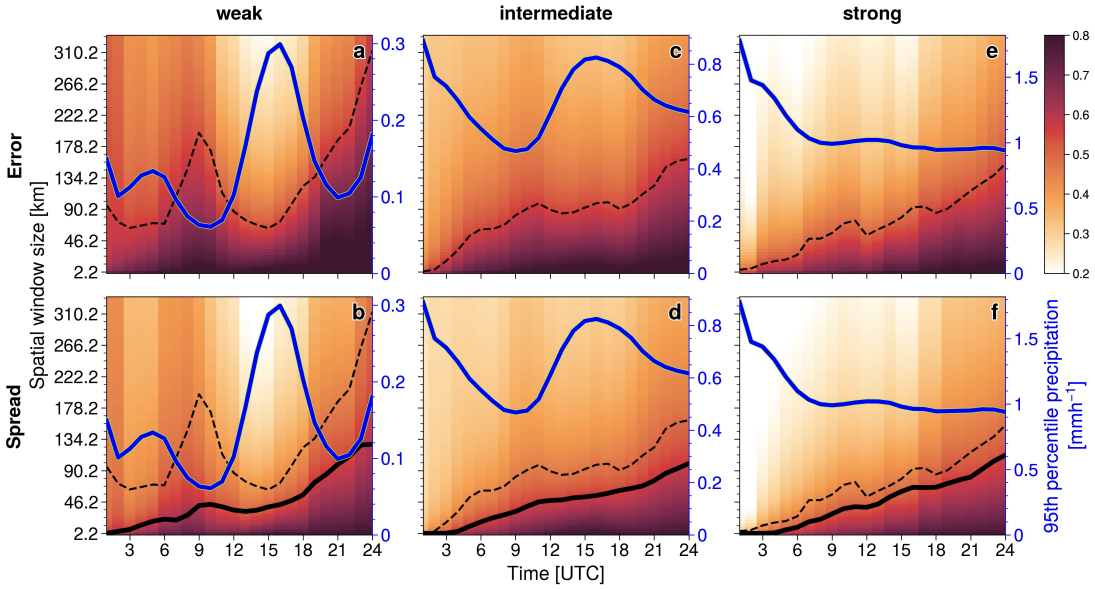

Figure 4.10: Weather regime-dependent, spatio-temporal variability of ensemble mean spatial error and spread, based on 20-member operational ICON-D2-EPS hourly precipitation forecasts, averaged over 16 weak forcing (a,b) and 16 strong forcing (e, f) cases. The centre column (c,d) illustrates the intermediate cases. The black dashed lines show median eDS of spatial error and the bold lines show median sDS representing spatial spread. The blue lines indicate the 95^{th} percentile values of hourly precipitation used in the FSS calculation.

Under strong convective forcing regimes, the composite precipitation time series shows a continuous decay of intensities throughout the day. The maximum precipitation intensity is six times larger (1.8 mm/h, Fig. 4.10e) than the maximum during non-equilibrium

conditions and occurs at 0000 UTC (as in Puh et al., 2023). This temporal evolution is dominated by a few strongly forced cases with nighttime maxima (e.g. the heavy precipitation causing the flooding in the Ahr valley on 14 July, not shown). The spatial error (eDS) increases almost continuously from model initialisation onwards (with the exception of slight dip around 1200 UTC) but stays at smaller spatial scales than in the non-equilibrium regime. Similar to the eDS, the sDS steadily increases, too (Fig. 4.10f). The proximity of spatial error and spread curves suggests a smaller spatial overconfidence compared to non-equilibrium. Moreover, the continual increase of spatial variability within the ensemble (measured by the sDS) caused by larger and larger displacements of rainfall patterns is a classic feature of error growth driven by scale interactions (Ying and Zhang, 2017).

For the intermediate-regime cases, time series of precipitation amount display a mixture between equilibrium and non-equilibrium categories in terms of amplitude and the diurnal cycle (Fig. 4.10c, d). The evolution of the DS for error (eDS) and spread (sDS) is more similar to the equilibrium situation, but the values of the spatial error and spread are systematically larger and closer to that of the non-equilibrium regime. This is not surprising, as roughly two third of summer convective precipitation over Germany occur in an equilibrium situation (Zimmer et al., 2011; Kühnlein et al., 2014). The weather-regime independent spatial error and spread largely parallels the intermediate category (not shown). Hence, only a flow-dependent examination of the evolution of spatial error and spread unveils the contrasting behaviour in non-equilibrium conditions.

Overall, the spatial forecast skill is systematically higher (i.e. smaller spatial error, compare dashed lines in Fig. 4.10a,c,e) indicating a superior forecast quality in equilibrium. Likewise, the spatial predictability is higher at most lead times (i.e. smaller spatial spread) during this particular flow pattern (compare bold lines in Fig. 4.10f). However, there is one important exception to this generalisation. During non-equilibrium, the spatial predictability is higher at the time of afternoon precipitation (lowest sDS) at forecast lead times of 12 to 18 hours (coinciding with time in UTC) that is strengthened by the orographic effect exerting a source of predictability (not shown). Enhanced uplift induced by orography promotes deep convection initiation and tends to organise convective cells along the orographic gradient, demonstrating a higher likelihood of convective activity in these regions. This structuring effect constrains the spatial variability of intense precipitation and increases the predictability. This effect has a greater impact in non-equilibrium regimes, where orographic triggering plays a more significant role, as observed over central Europe (Bachmann et al., 2020; Keil et al., 2020) and the British Isles (Flack et al., 2018), for instance.

4.3.3 Systematic impact of PSP

Building upon the flow-dependent assessment of the operational forecast system, we proceed to quantify the impact of incorporating a novel model uncertainty representation, specifically the PSP scheme in the ICON-D2 ensemble. Usage of the PSP scheme shows a large systematic impact on the diurnal cycle of precipitation in the weak convective for-

cing regime (TP ensemble). Precipitation starts earlier, peak precipitation rates become stronger and precipitation decays slightly earlier thereafter (Fig. 4.11a, b). Note that the total daily rainfall is hardly affected by the PSP scheme (± 3 % relative difference). The impact on spatial error and spread is discernible from 0900 UTC onwards, once convection is initiated. For most of the time (0900–1900 UTC, including the convective most active period) the PSP scheme reduces spatial forecast errors of precipitation at scales larger than 20 km (Fig. 4.11a). The spatial error increase after 1900 UTC is presumably linked to an earlier decay of precipitation using the PSP scheme, a known issue applying the scheme (Rasp et al., 2018). Despite this earlier reduction in precipitation intensity the general precipitation structure persists leading to the time lag between the decrease in precipitation and the deterioration of spatial error.

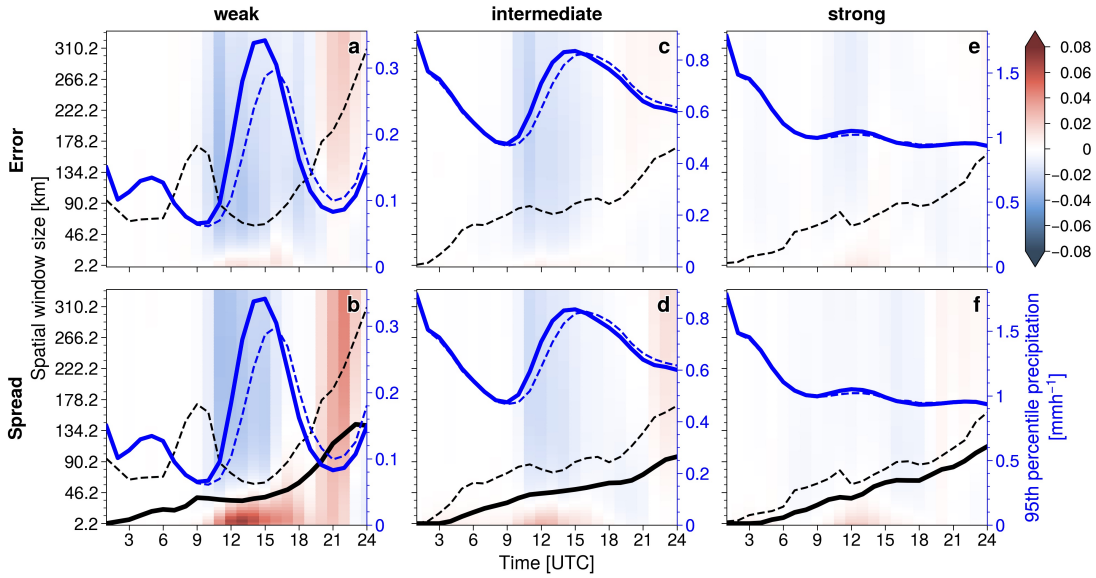

Figure 4.11: Same as Fig. 4.10, but the composite of FSS differences of TP minus TR ensembles. The blue lines indicate the 95^{th} percentile values of hourly precipitation of TP (bold) and TR (dashed blue) forecasts. The black dashed (bold) line shows the median eDS (sDS) of the TP simulation.

In general, and similar to the spatial error, the impact of PSP on spatial spread shows a distinct spread decrease at scales larger than roughly 50 km between 0900–1700 UTC (Fig. 4.11b). Strikingly there is a steady increase of spatial spread on scales smaller than 50 km in this time window. By design the PSP scheme inserts perturbations at the effective model resolution (about 10km) which consequently cause increased variability in the boundary layer impacting convective processes among others. Due to upscale error growth, this affects larger and larger scales as time goes by and leads to a continual spatial spread increase of hourly precipitation rates at spatial scales rising from the model's effective resolution to some tens of kilometres. The reduction of spatial error and spread at scales larger than 50 km can be attributed to the strong penalty associated with complete misses

in forecasts when calculating the FSS. By fostering convection initiation in members of the TP ensemble that failed to predict convection in the TR ensemble, the number of complete misses is diminished and the FSS based spatial error and spread is reduced across scales. Vice versa, this effect causes an increase in spatial error and spread across scales when convection ceases earlier in the evening using the PSP scheme.

For the other forcing categories the PSP scheme shows characteristics similar to the weak forcing (but at reduced levels). Figure 4.11b, d, f indicates that the spread shows a similar behaviour across all regimes but at successively smaller amplitudes for stronger forcing conditions.

4.4 Summary and Conclusions

The spatial predictability of precipitation strongly depends on the prevailing convective forcing regime. For a systematic assessment, we condense three months of operational ICON-D2-EPS forecasts into varying convective forcing categories applying the convective adjustment time scale diagnostic. The scale-dependent aspect is assessed using spatial error (eDS) and spread (sDS) based on the FSS technique. During weak convective forcing, corresponding to the non-equilibrium regime, the spatial error and spread largely depends on the diurnal cycle of precipitation. The median eDS is about 70 km during the convectively most active period (1200-1800 UTC) whereas the sDS amounts to only 40 km. The scales between 40 and 70 km show no spatial skill and are under-dispersive. During strong convective forcing, equivalent with the equilibrium regime, the large-scale flow constrains convection patterns stronger and the gap of the median eDS and sDS becomes smaller. The spatial forecast quality is superior in this weather regime, with the eDS fairly continuously increasing from model initialisation. The lack of spatial spread suggests that the current perturbations in ICON-D2-EPS are sub-optimal and additional sources of model uncertainty need to be added.

The application of the PSP scheme systematically shows the anticipated beneficial impact. Whereas the total daily rainfall is hardly affected by the PSP scheme, it remarkably reduces spatial forecast errors of precipitation at scales larger than 20 km, in particular in the non-equilibrium regime during the diurnal cycle from 0900 UTC onwards. Whereas the spatial spread is increased by PSP at scales less than about 50 km, the variability is decreased at larger scales. For the other forcing categories the PSP scheme shows similar characteristics but at successively smaller amplitudes for stronger forcing situations. The PSP scheme is effectively representing subgrid-scale uncertainty in the boundary-layer turbulence and reduces the systematic error as aspired by stochastic parameterisations (Berner et al., 2017).

The effect of additionally including microphysical parameter perturbations is explored for two representative weather situations by constructing prototype 'Grand Ensemble' ICON-D2 experiments. Univariate and multivariate IBC, PSP and MPP ensemble simulations for representative cases allow us to disentangle individual and synergistic contributions of the sources of uncertainty. They confirm the strong flow-dependent impact of

PSP, which has a significant impact on ensemble and spatial spread in non-equilibrium, while its effects are negligible in equilibrium. The time and scale of the maximum difference is at the onset of convection and at the scale of the perturbations, which is beneficial, because the operational ICON-D2-EPS struggles to introduce sufficient spatial variability in the precipitation field at that time. In comparison, IBC represent the primary source of variability in the ensemble in equilibrium. The joint impact of PSP and MPP in the presence of IBC uncertainty regarding spatial error and spread is additive, with PSP being the primary model uncertainty in non-equilibrium. Precipitation is triggered earlier and its intensity is higher. The spatial error is largely reduced at convection initiation and the spatial spread is increased across time and scales. Hence the spatial error-spread relationship is significantly improved compared to the pure IBC ensemble. The conclusions on the additive impact of MPP hold for these two cases and for this ensemble configuration. However, longer term testing of the synergistic impact of model uncertainties is required to achieve more robust results.

In summary, ICON-D2-EPS forecasts show insufficient spatial spread of precipitation compared to spatial forecast errors. This agrees with earlier findings for other CPEPSs of Rezacova et al. (2009) and Dey et al. (2014). Stratifying this shortcoming by convective forcing regimes reveals that the spatial under-dispersion emerges especially in non-equilibrium conditions. The application of physically-based stochastic perturbations in the planetary boundary layer considerably reduces the spatial overconfidence. This paper emphasises the importance of the flow-dependent approach and supports previous results of variable predictability of area-averaged precipitation amounts (Keil et al., 2019; Barthlott et al., 2022b; Matsunobu et al., 2022) and scale-dependent results (Flack et al., 2018, 2019; Matsunobu et al., 2022). Tracing the varying sensitivity of uncertainty representations to different origins using a 'Grand Ensemble' with many formulations of model uncertainty will be pursued in future work and is presently left as a challenging open question.

Chapter 5

Convergence of impact and interactions of model uncertainties

The previous chapters have demonstrated how various sources of uncertainty act on different amplitudes, times, and scales. As discussed in Sect. 1, further assessments of the model uncertainty impact are warranted to identify model uncertainty representations that are effective in improving ensemble variability. To provide a handy framework for these assessment, a variability budget method is proposed (Sect. 2.5). This method enables to extract model uncertainty variances from total variance and thus to quantify a sensitivity of ensemble variability to the model uncertainty. Such a sensitivity analysis in general suffers from sampling error. On the other hand, utilising this method allows measuring convergence of the model uncertainty variance, which is informative for optimising an ensemble design, but has not been investigated in the context of NWP.

Motivated by this idea, this chapter investigates the minimum level of randomness required for accurate estimation of model uncertainty impact, addressing the question **Does model uncertainty need as many ensemble size as IBC to accurately estimate variability?** The convergence behaviour of variance and correlations introduced by the PSP scheme is examined with respect to the number of IBC (nIBC) and random seeds (nRSD) in the PSP scheme. Utilising bootstrapping, directly sampled convergence behaviour is compared with the scaling theory of sampling error for standard deviations (Craig et al., 2022; Tempest et al., 2023) and correlation coefficients (Fisher, 1921).

5.1 Ensemble experiment and convergence measure

5.1.1 Ensemble dataset

The ensemble dataset was constructed as follows. We picked out one representative weak-forcing case, 10 June 2021 as a case study. Initially, a 20-member ensemble was formed with perturbations only in IBC, named the I ensemble. Next, for each member of the I ensemble, 20 additional simulations with different RSDs in the PSP scheme were performed,

resulting in the $I \times P$ ensemble comprising $20 \times 20 = 400$ members.

Bootstrapping (resampling with replacement) was used to assess the dependence of variance on nIBC as well as nRSD. For the I ensemble, sub-ensembles of size s were randomly drawn from the original 20 members 100 times, denoted as I_s, where s is successively increased from 1 to 20. Similarly, the $I \times P$ ensemble underwent bootstrapping considering both nIBC and nRSD. By randomly selecting s IBCs and t RSDs for each IBC, sub-ensembles of size $s \times t$ were created, named $I_s \times P_t$. Impact of the PSP scheme is derived by comparing RSD-perturbed members in $I_s \times P_t$ with the corresponding members in I_s sharing the same IBCs. The derived sub-ensemble of the PSP impact is denoted as $P_t|I_s$ and depends on both nIBC (s) and nRSD (t). Finally, correlation coefficients between IBC and PSP impact for any combination of nIBC and nRSD were calculated using Eq. 2.13 (omitting the perturbation symbol x' for convenience):

$$Corr(I_s, P_t|I_s) = \frac{Var(I_s \times P_t) - Var(I_s) - Var(P_t|I_s)}{2\sqrt{Var(I_s) * Var(P_t|I_s)}}. \tag{5.1}$$

5.1.2 Convergence measure

To understand how many ensemble members are required to obtain stable variance, we analyse the saturation of variance values and confidence intervals (CI) of variance as a measure of sampling error. We calculate variance for each bootstrapping repetition and use the mean variance across repetitions as the most probable variance for a specific combination of nIBC and nRSD. We use the standard deviation of variances across repetitions as a measure of sampling uncertainty for variance. Given that the sampling distribution of the summation of the random variables is expected to be normally distributed (Dekking et al., 2005), the standard deviation corresponds to a half of the width of a 68.3 % CI.

When calculating CIs for correlation coefficient, a careful treatment is needed due to their non-symmetrical nature and the 0-1 boundary. We employ a bootstrapping method similar to that used for variance, which calculates the correlation coefficient for each repetition with the 68.3 % CI derived from the difference between the upper (84.13 %) and lower (15.87 %) percentile values of the sampling distribution. This approach avoids assumptions about the distribution shape and allows us to account for the potentially unequal influence of nIBC and nRSD.

It is known that sampling uncertainty of variance (standard deviation) decreases proportionally to $n^{-1/2}$ with an increase in sample size (Craig et al., 2022; Tempest et al., 2023). On the other hand, the estimation of sampling uncertainty for correlation coefficients needs a bespoke approach due to their asymmetry. It is convenient if there is an available benchmark as to variance to assess if sampling uncertainty of correlation can be extrapolated to larger ensemble size. One method to parametrically obtain an asymmetric CI for correlation coefficients utilises Fisher's r-to-z transformation (Fisher, 1921), which transforms a sample correlation coefficient r into a z-score z through a logarithmic function:

$$z = 0.5 \log \frac{1+r}{1-r}. \tag{5.2}$$

The transformed z-score assumes an approximately normal distribution with a known standard deviation of $1/\sqrt{n-3}$ (Fouladi and Steiger, 2008). An expectation of z with sufficient sample size is expected to be equal to the transformation of the true correlation coefficient ρ. Thus, its z-score is $z_{true} = 0.5 \log \frac{1+\rho}{1-\rho}$. The distribution of z can be normalised as:

$$\frac{z - z_{true}}{1/\sqrt{n-3}}. \tag{5.3}$$

Normalisation allows us to easily obtain quantile values of the sample distribution using a look-up table, thereby enabling the calculation of the upper and lower bound of the α% CI of z_{true}:

$$-z_{\frac{100-\alpha}{2}} \leq \frac{z - z_{true}}{1/\sqrt{n-3}} \leq z_{\frac{100-\alpha}{2}}. \tag{5.4}$$

And

$$z - z_{\frac{100-\alpha}{2}} \frac{1}{\sqrt{n-3}} \leq z_{true} \leq z + z_{\frac{100-\alpha}{2}} \frac{1}{\sqrt{n-3}}. \tag{5.5}$$

By applying inverse transformation of Eq. 5.2 to the calculated CI in z-space, one can derive the upper and lower values of ρ:

$$\rho = \frac{e^{2z} - 1}{e^{2z} + 1}. \tag{5.6}$$

Fig. 5.1 illustrates how a Fisher-z' value varies with the input correlation coefficient for sample size of 20. As the correlation coefficient increases, the z values exhibit a monotonic increase, with the rate of change reaching maximum near the correlation coefficient of -1 and 1, and its minimum around r = 0. When correlation coefficient indicates no correlation, even a small change in the z value can result in a substantial change in the correlation coefficient. A fixed width of the z value specified in Eq. 5.5 results in smaller CI widths for large correlation coefficients and, in contrast, larger CI width for small correlation coefficients, as depicted in the right panel of Fig. 5.1.

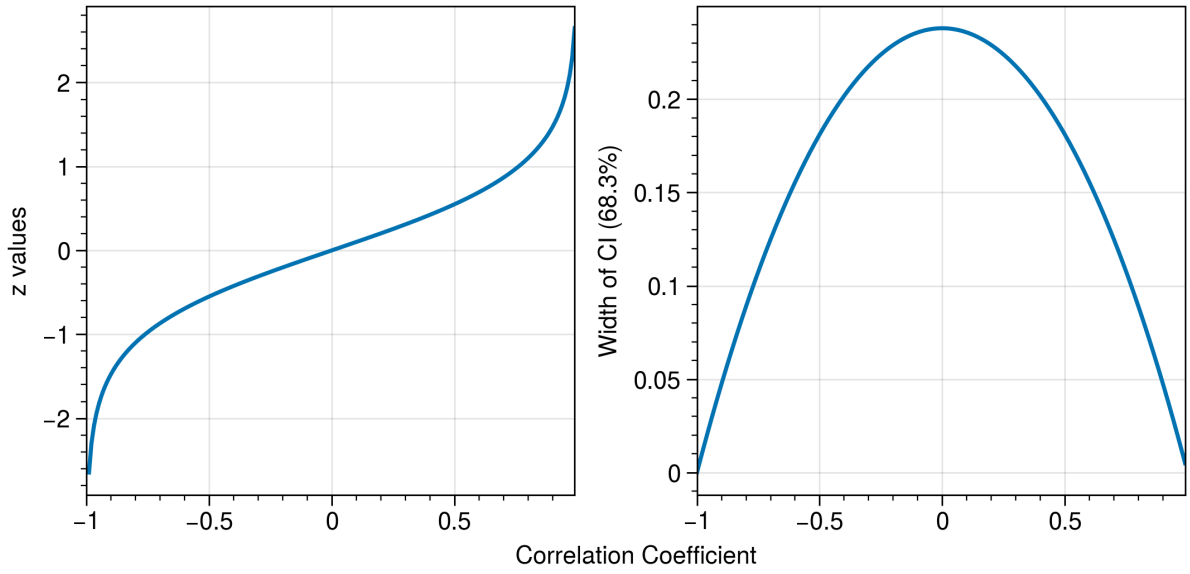

Figure 5.1: Example Fisher's z values and a half width of 68.3 % CI as a function of input correlation coefficient.

To assess whether convergence of correlation coefficients obtained through bootstrapping aligns with the theoretically expected convergence rate, we compare the width of the 68.3% CI between the bootstrapping and Fisher-z transformation methods. In the Fisher-z estimation, the true correlation coefficient ρ is estimated from the repetition mean values of the bootstrapped correlation coefficients. Note that since a 68.3% CI corresponds to twice a standard deviation for variance sampling error, we present half the CI width to facilitate easier interpretation of results.

5.2 Temporal evolution of variance

Since the focus of this study is on sampling uncertainty of variability attained by the PSP scheme, we should assess variables influenced by the scheme at lead times when the scheme could be effective. To identify such variables and forecast times, the temporal evolution of ensemble variances are shown in Fig. 5.2. The temporal patterns of variance are strongly height-dependent, with upper-level variance (blue lines) increasing steadily with lead time and lower levels (pink, green) exhibiting diurnal cycles peaking in the early afternoon, for all variables except for geopotential. The 500 hPa level (orange) displays the patterns like a mixture of the lower and upper levels except for vertical wind, where convective updraft appears to be the primary source of variability and thus strongly associated with the activity in the lower troposphere. This reflects the dominance of large-scale flows at upper levels e.g. geostrophic balance and the strong influence of diurnal convective activity at lower levels.

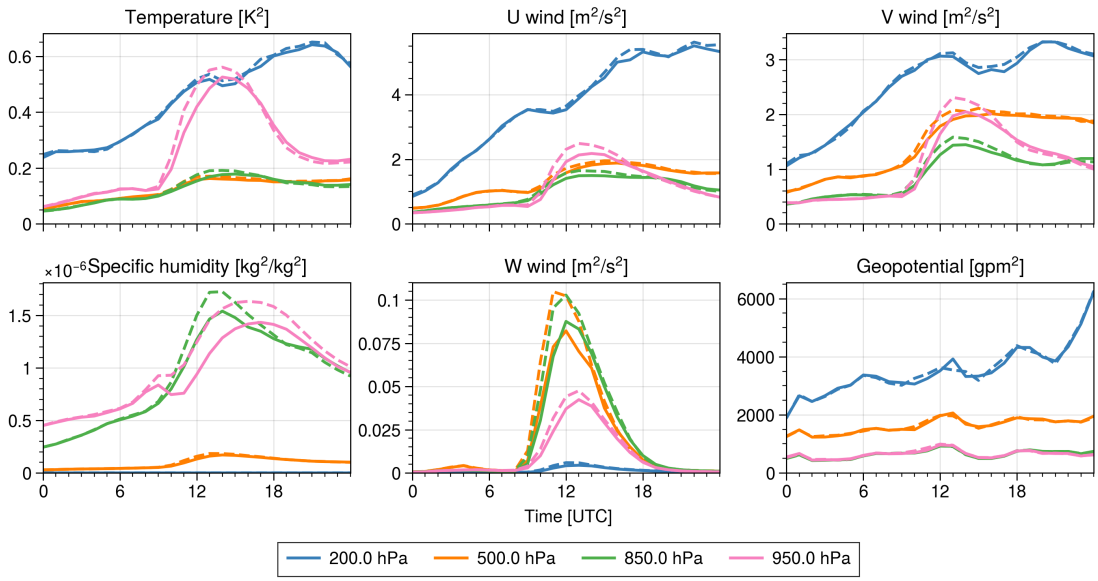

Figure 5.2: Time evolution of ensemble variance of the I_{20} (solid) and $I_{20} \times P_{20}$ (dashed) ensembles averaged over the German domain.

Comparing solid and dashed lines in Fig. 5.2 highlights five key points regarding the impact of the PSP scheme on variance: 1) Lower levels (950 and 850 hPa) for temperature, winds and humidity are generally sensitive to PSP as well as 500 hPa vertical winds which is particularly sensitive. This is no surprise considering that those variables are strongly disturbed by the convective activity which the PSP scheme is supposed to impact. 2) In contrast, upper tropospheric levels (200 and 500 hPa) are insensitive. This is expected given the absence of synoptic-scale vertical motion over Germany on that day; 3) PSP primarily increases variance associated with the diurnal cycle between 1000 and 1800 UTC; 4) Geopotential fields show little sensitivity to PSP, indicating overall insensitivity to local convection; and 5) Overall, the net increases in variance given by PSP remain relatively small compared to the variance of I. This small net effect of adding model uncertainty representations on total variance aligns with existing literature on model uncertainty representations (e.g. Buizza et al., 1999; Berner et al., 2017; Flack et al., 2021; Hermoso et al., 2021; Frogner et al., 2022; Puh et al., 2023).

The results in the previous paragraph strongly suggest the investigation should be guided by the diurnal cycle of the convective activity. For further examination of how well the additional uncertainty are represented at different stages of convection, we examine four specific time steps throughout the forecast: 0800 UTC (pre-convection), 1100 UTC (convection onset), 1400 UTC (peak convection), and 2000 UTC (post-convection). Figure 5.3 shows the distribution of accumulated precipitation at these lead times for later comparisons with variance distributions. Precipitation along the Alps occurs consistently throughout the forecast, while strong localised convection appears in southern Germany from 1100 UTC onwards. No convective precipitation is observed in northwestern Ger-

many. In selecting variables for the analysis, we don't include geopotential and vertical wind due to its insensitivity to convection and lack of variance outside the convective time period, respectively. U wind and V wind are expected to provide similar results. Thus, we focus on temperature at the 850 hPa level, as it is in the levels sensitive to the scheme and likely located above the planetary boundary layer, where tendencies are directly perturbed. To provide a comprehensive result, we also present some results for u-wind and specific humidity at the 850 hPa level.

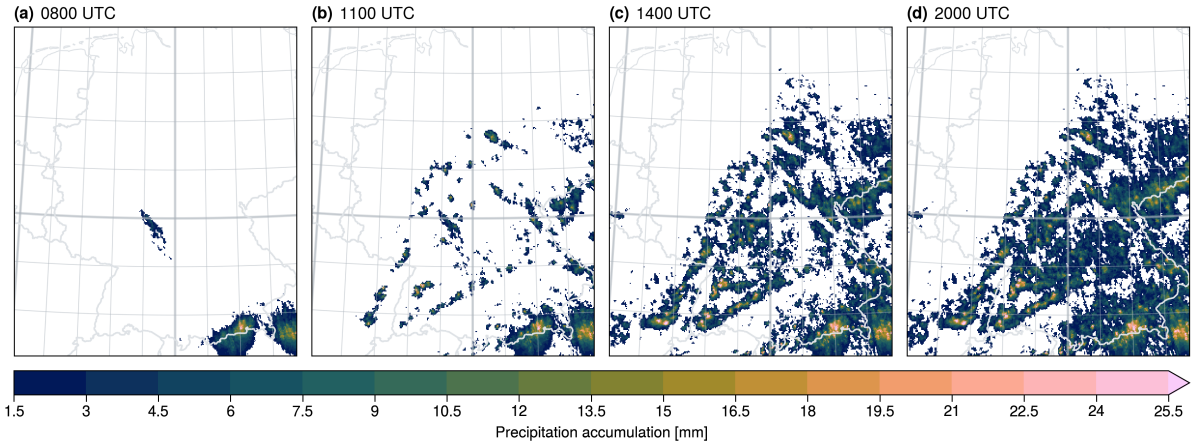

Figure 5.3: Precipitation accumulation from 0000 UTC. The ensemble mean of the I_20 ensemble is displayed.

5.3 Variance convergence for different perturbations

Figure 5.4 shows the evolution of mean variance and sampling uncertainty before convection with an increase in nIBC and nRSD. The variance of I_s and $I_s \times P_{20}$ largely overlap and rapidly increase as nIBC increases from 1 to 4, and afterwards slowly increase and become almost saturated with nIBC of 15 (Fig. 5.4a). The variance of $P_{20}|I_s$ remains low and has negligible influence on the total variance $I_s \times P_{20}$. Both variance of $P_t|I_{20}$ and $I_{20} \times P_t$ show no sensitivity to nRSD (Fig. 5.4b), indicating the PSP scheme's inactivity before convection.

Sampling uncertainty, as measured by the CI width, continuously decreases for the variances of I_s and $I_s \times P_{20}$ with increasing ensemble size (Fig. 5.4c). This decline is proportional to $nIBC^{-1/2}$, consistent with the findings by Craig et al. (2022) and Tempest et al. (2023). On the other hand, nRSD has no effect on the sampling uncertainty (Fig. 5.4d). Increasing nRSD at this time is like making copies of identical nIBC-member sub-ensembles, leading to constant variance and sampling error.

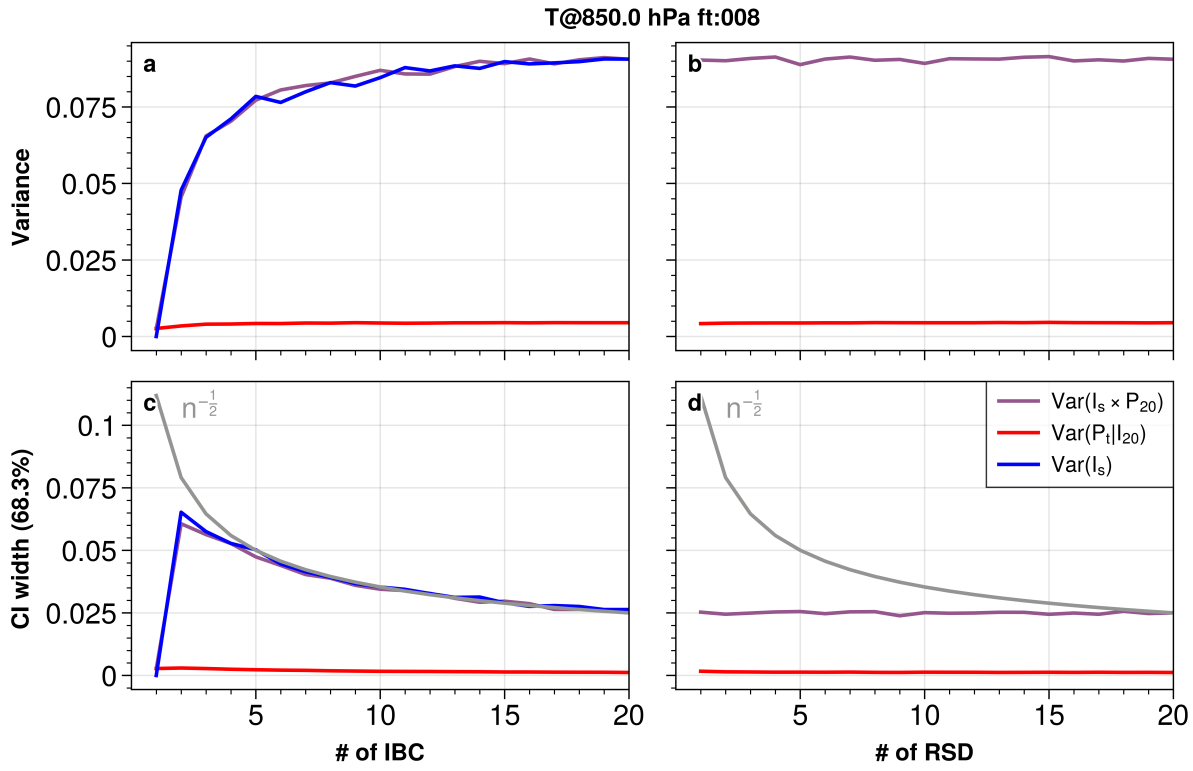

Figure 5.4: Area-averaged variance (upper) and sampling uncertainty (lower) as a function of nIBC (left) and nRSD (right) of temperature at 850 hPa pressure level before convection (0800 UTC).

At the time of convection initiation (1100 UTC), the PSP scheme starts to act (Fig. 5.5a). With only 1 IBC, the ensemble variability is solely attributed to the PSP scheme, resulting in comparable variances for $I_s \times P_{20}$ and $P_{20}|I_s$. Similar to the pre-convection case, the variance of $I_s \times P_{20}$ increases rapidly and attains values exceeding the I_s variance by roughly 10%. The variances of $P_{20}|I_s$ also increases as nIBC gets larger. This can be explained by the PSP scheme's influences on specific regions, where the atmospheric state needs to be potentially uncertain, and multiple IBCs allow the PSP scheme to be active on different grid points due to the difference in the initial state. However, its growth is slower than the I_s variance and saturates earlier with 6 nIBC at a lower level. Notably, nIBC of 15 remains a suitable sample size for estimating variances averaged over Germany, regardless of the variance type. This number is only valid for a domain average. A grid-point variance needs much more members to get converged because taking a domain average increases effective sample size, as shown in Craig et al. (2022). While nRSD's impact becomes visible, it plateaus sooner compared to nIBC (Fig. 5.5b). With only one RSD, both $P_t|I_{20}$ and $I_{20} \times P_t$ variances attain nearly 90% of the variance obtained with 20 RSDs. Subjectively, using more than 3 RSDs offers no further benefit to variances.

The sampling uncertainty of the I_s, $I_s \times P_{20}$, and $P_{20}|I_s$ variances all decrease propor-

tionally to nIBC to the power of $-1/2$ (Fig. 5.5c). Whilst the smaller sampling uncertainty for $P_{20}|I_s$ likely arises from its smaller variance values, the $I_s \times P_{20}$ variance shows smaller sampling uncertainty than I_s despite having a larger variance. The PSP scheme activation likely enlarges the effective ensemble size of $I_s \times P_{20}$. Interestingly, increasing nRSD modestly reduces the sampling uncertainty, at a slower rate than the $n^{-1/2}$ line (Fig. 5.5d). If the impact of the PSP scheme exhibits perfect randomness, a sub-ensemble $I_s \times P_t$ has a normally distributed sample distribution, and its sampling uncertainty reduces according to the central limit theorem, i.e. $nRSD^{-1/2}$. This can be because, with 20 IBCs, the IBC uncertainty to large extent covers a potential range of the PSP uncertainty, and further increase in nRSD only adds insignificant information about the sample distribution.

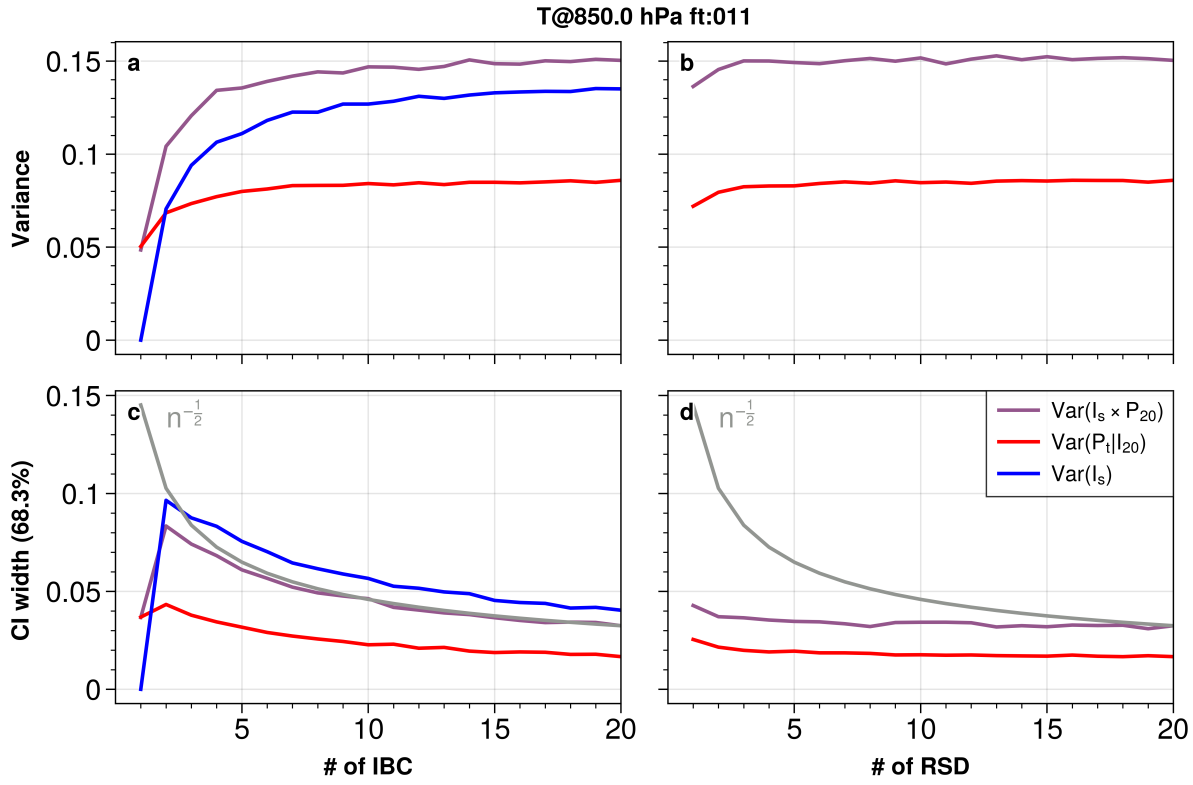

Figure 5.5: As for Fig. 5.4, but at the time of convection initiation (1100 UTC).

The temperature variances at peak convection (1400 UTC) resemble those at 1100 UTC, with a larger variance of $P_t|I_s$ (Fig. 5.6, left column). After convection (2000 UTC), the variances of I_s and $I_s \times P_t$ largely overlap, indicating that I_s plays the dominant role in determining the variance of $I_s \times P_t$ similar to the pre-convection phase (Fig. 5.6, right column). Whilst the variance of $P_t|I_s$ remains positive, it becomes substantially smaller than the other variances. This suggests that the PSP's impact persists in the model atmosphere even after convection has ceased, but it is negatively correlated with the IBCs and does not contribute to the overall variance increase in $I_s \times P_t$ (physical and statistical

interpretations of this correlated impact are provided in Chapter 6). The sampling uncertainty of the $P_t|I_s$ ensemble continues to decrease proportionally to $nIBC^{-1/2}$, similar to the behaviour during convective periods. It appears that the presence of preceding convection, rather than being in a specific stage of convective events, is the primary factor governing the PSP variance. While larger nRSD contributes to a decrease in sampling error once convection occurs, the rate of decrease is considerably slower and overall less effective compared to nIBC.

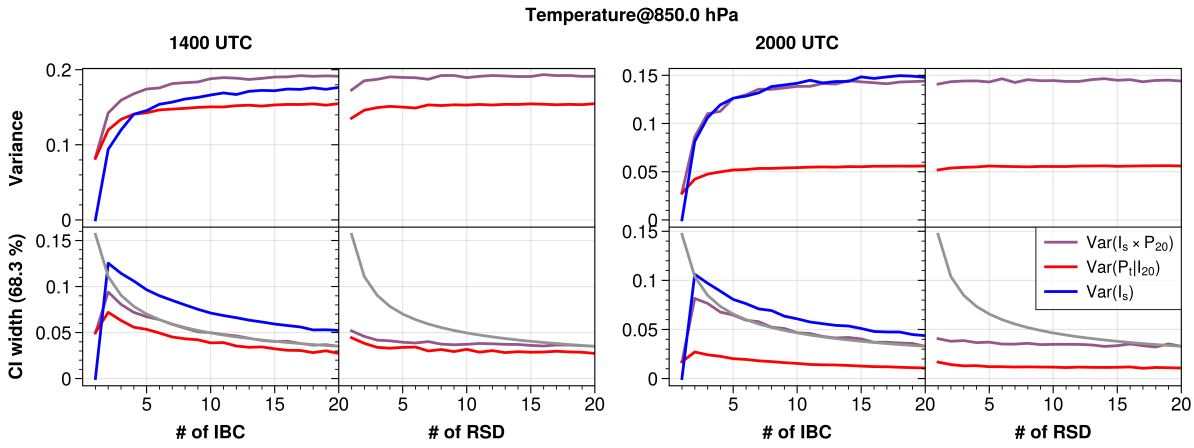

Figure 5.6: As for Fig. 5.4 and 5.5, but at the peak time of convection (1400 UTC; left) and after convection (2000 UTC; right).

For U component of wind, the variance of $P_t|I_s$ (red solid) exhibits large values, indicating stronger sensitivity to the PSP scheme (upper in Fig. 5.7). The curves of the $P_t|I_s$ variance parallel the I_s variance (blue solid) and appear to saturate around nIBC = 10, which is larger than the case for temperature (nIBC = 6). This indicates that necessary nIBC for variance saturation depends on variables, such as dynamics rapidly smooth temperature fields and lead to quick saturation of the $P_t|I_s$ variance for temperature above the boundary layer, but not as for U wind. Helped by the increase in effective sample size by the PSP scheme, the variance of $I_s \times P_t$ (red dash) shows the smallest sampling uncertainty and converges around 0.35 when nIBC = 20. Despite the distinct sensitivity of U wind variance to nIBC compared to temperature, the sensitivity to nRSD is similar.

For nRSD, three members seems sufficient for adequate sampling of the PSP variance. For sampling uncertainty, the convergence curve of $I_s \times P_t$ (red dashed) starts slightly above 0.5 and shows steeper decrease than that for temperature with small nRSD, converging around 0.35 with nRSD of 10 and larger. Variables responsive to the PSP scheme require several random seeds to stabilise variance. Nevertheless, the convergence rate is smaller than $nRSD^{-1/2}$ and smaller nRSD does not significantly impair variance sampling. This conclusion holds for specific humidity, the most sensitive variable to the PSP scheme (lower in Fig. 5.7).

Former analyses show the dependency of the impact of PSP on the IBC impact on

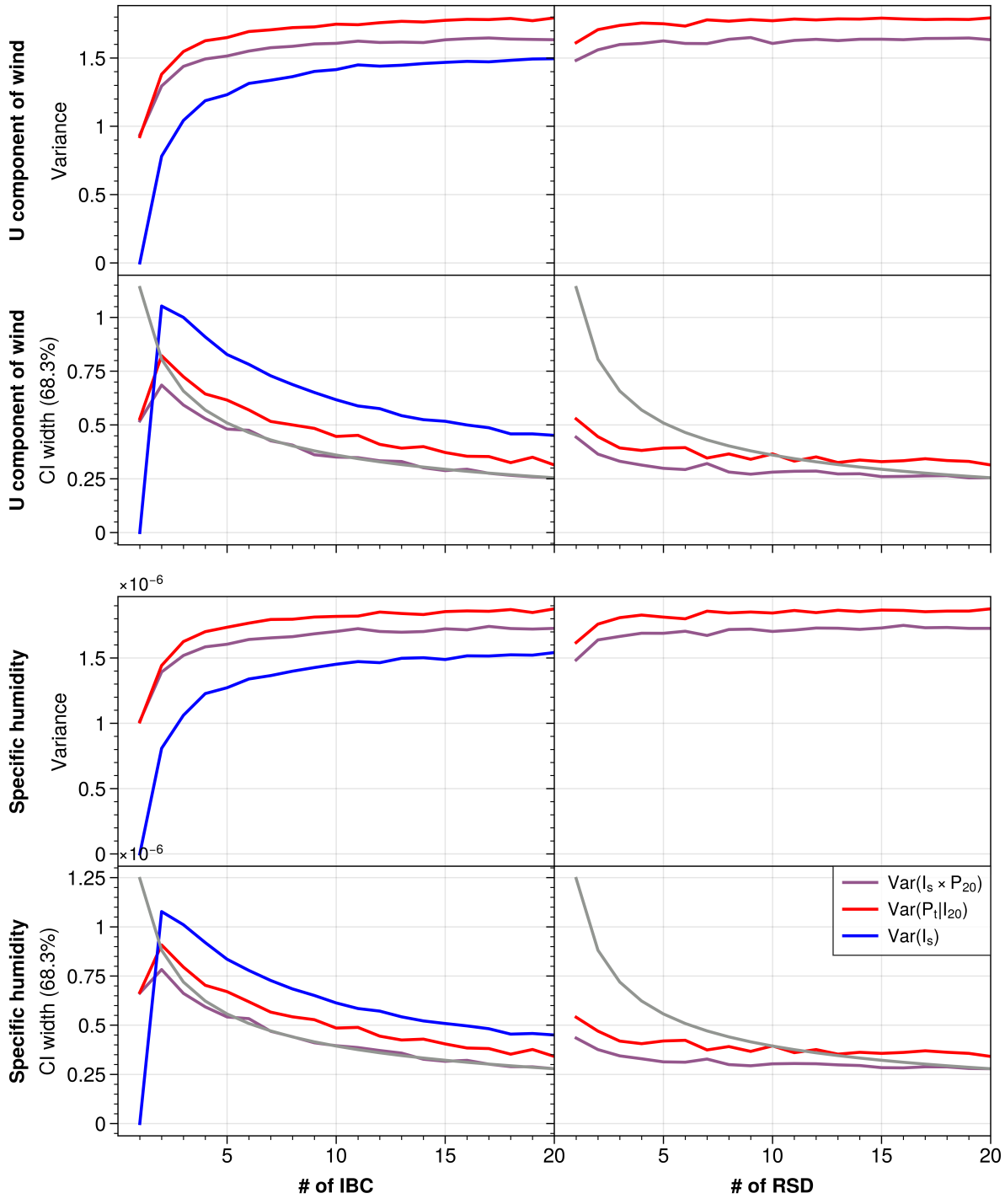

Figure 5.7: As for left column in Fig. 5.6 but for u component of wind and specific humidity at the peak time of convection (1400 UTC).

variance convergence. To compare the spatial distribution of the PSP impact with IBC, the spatial pattern of the $I_s \times P_t$ variance as a function of nIBC and nRSD at peak convection (1400 UTC) is shown in Fig. 5.8. As expected, nIBC plays a more significant role than nRSD in attaining larger variance. Increasing nIBC from 2 to 6 leads to an area-averaged variance increase of approximately 50%, while increasing nRSD from 1 to 20 yields a gain of 40% for nIBC=2 and 10% for nIBC $\geq$ 10.

The variance bands in northern Germany and over the Alps exhibit distinct changes with varying nIBC and nRSD. In these regions, increasing nIBC primarily intensifies the variance, while increasing nRSD broadens the spatial extent of the variable region. In contrast, a careful comparison of plots within each row reveals that nRSD can structure small-scale variability in southern Germany, apparently related to ongoing small-scale convective activity in that region (Fig. 5.3c). As expected from its design, the additional randomness introduced by the stochastic fields in the PSP scheme selectively enhances small-scale variability through local convection in a way that increasing nIBC cannot achieve. This small-scale impact of model uncertainty representation is consistent with the recent work by Hermoso et al. (2021) and Chen et al. (2024), with the latter showing that the model uncertainty impact saturates at the 100 km scale in 48 hours. However, nRSD's impact on the spatially averaged variance remains small, only around 10% increase, when nIBC is sufficiently large (> 10).

5.4 Effect of sampling uncertainty on estimating correlation

This subsection examines how convergence and sampling uncertainty in the variance influence the estimation of correlation coefficients. Figure 5.9 depicts the spatial distribution of variances and the correlation coefficient associated with the impact of IBC and PSP. Comparing the variance patterns in Figs. 5.9a, b, and c reveals that adding PSP weakens the magnitude of the I_{20} variance signal in northern Germany, consistent with the blurring effect discussed in the previous section. In southern Germany, the variance of $I_{20} \times P_{20}$ exhibits a relatively smoothed pattern, whereas the I_{20} and $P_{20}|I_{20}$ variances show more intermittent signals, with a broader distribution of weak variance for $P_{20}|I_{20}$. The PSP scheme enhances variability in regions where the atmospheric state is potentially uncertain in I_{20}, as well as enhancing small-scale variability missed by the IBC uncertainty, resulting in a smoothed variance field for $I_{20} \times P_{20}$. Regarding the correlation (Fig. 5.3d), a negative signal is evident wherever both $P_{20}|I_{20}$ and I_{20} exhibit some extent of variance, particularly extending from southwest to northeast Germany. This pattern closely resembles the precipitation map in Fig. 5.3c. The analysis confirms a negative correlation between the impacts of IBC and PSP, which is particularly pronounced in regions experiencing convective precipitation.

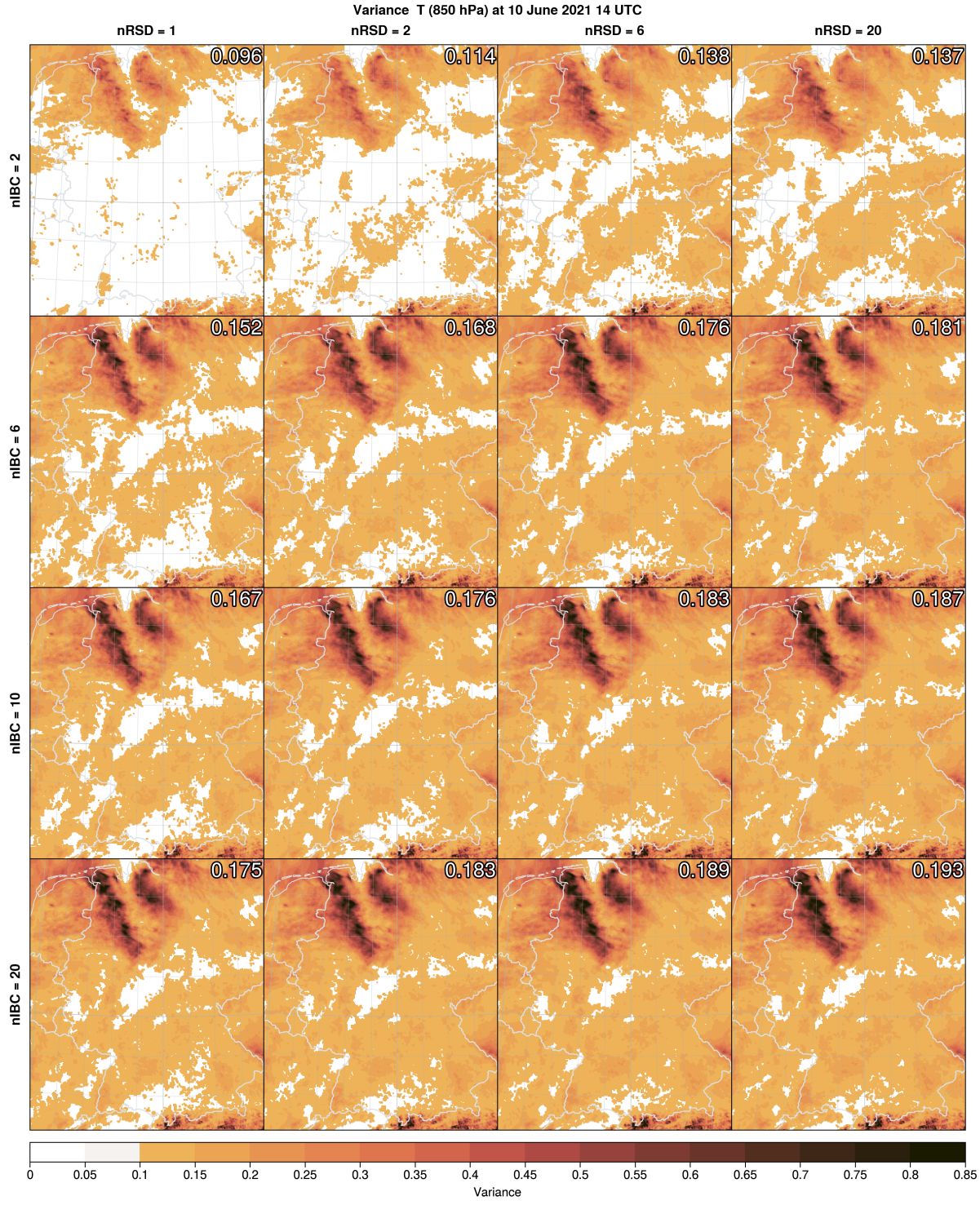

Figure 5.8: Total temperature variance [K^2] map at 850 hPa at the peak convection as a function of nIBC (rows) and nRSD (columns). Area-averaged variances are shown at the top-right corner.

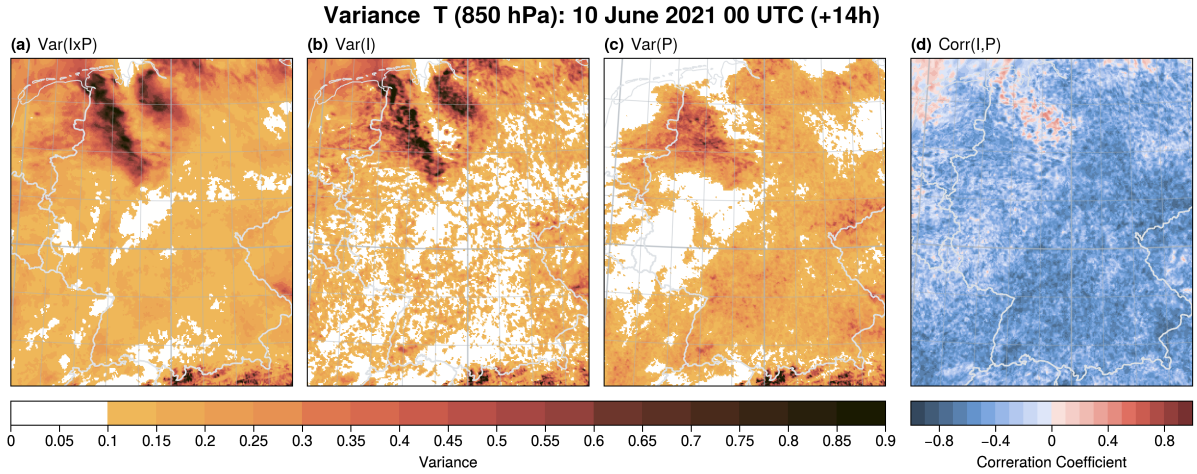

Figure 5.9: Spatial distributions of iteration-mean variance of (a) $I_{20} \times P_{20}$, (b) I_{20}, (c) $P_{20}|I_{20}$, and (d) correlation coefficient for 850 hPa temperature at mid convection (1400 UTC).

Figure 5.10 shows the variation of sampling error in the correlation coefficient with nIBC and nRSD. As evident from the area-averaged sampling errors, an increase in both nIBC and nRSD significantly reduces the sampling uncertainty. Increasing nIBC from 2 to 20 leads to a fourfold reduction in sampling uncertainty, while increasing nRSD from 1 to 20 yields roughly half that reduction. Similarly to the variance increase in Fig. 5.5, the impact of larger nIBC and nRSD on reducing sampling uncertainty is more pronounced for smaller initial values, and diminishes for nIBC and nRSD greater than 10. For example, increasing nIBC from 2 to 6 reduces uncertainty to 40%, while increasing from 10 to 20 only yields 24% reduction. Regions with high correlation coefficients in Fig. 5.9d, such as southern Germany, experience faster convergence with both nIBC and nRSD. Conversely, regions exhibiting noisy correlation patterns, like northwest Germany, suffer from persistent sampling uncertainty even with the largest ensemble size. This necessary large sample size for weak correlation aligns well with the expectation from the Fisher-z transform (Fig. 5.1). Comparison of average sampling uncertainty between the same total sample size but with different combinations of nIBC and nRSD shows systematically smaller uncertainty for combinations with larger nIBC, except the combinations of (20, 1) and (10, 2). Having more than one RSD is significant in reducing sampling uncertainty, especially in regions with weak correlation (northwest in Fig. 5.9d). Thus, although increasing nRSD in order to reduce correlation sampling uncertainty is less effective than increasing nIBC, it is worth considering when the current nRSD is just one. This highlights the importance of both NIBC and nRSD in reducing sampling uncertainty, especially in regions with weak correlations.

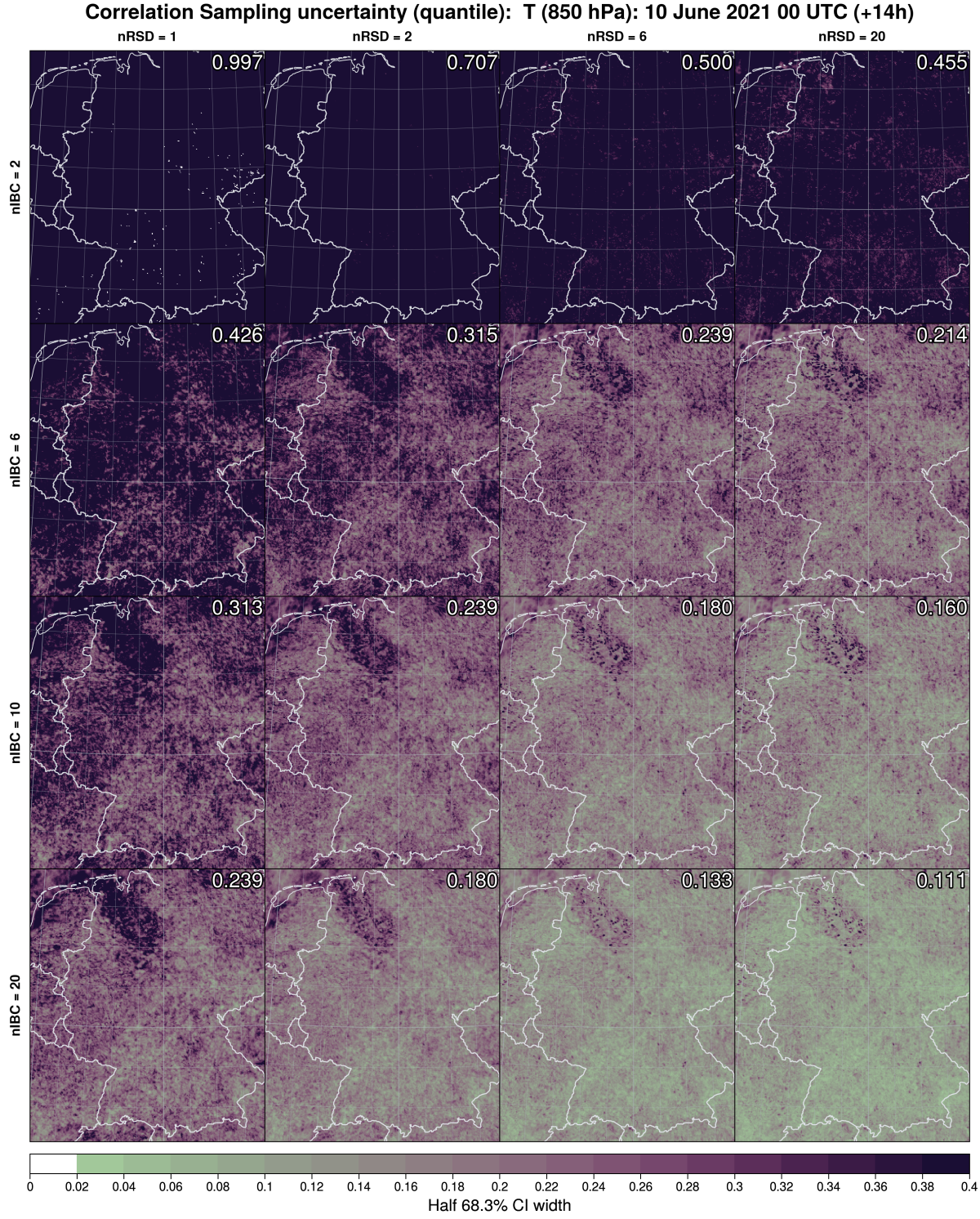

Figure 5.10: Sampling uncertainty map of the IBC-PSP correlation coefficient at 850 hPa at the peak convection as a function of nIBC (rows) and nRSD (columns). Area-averaged values are shown at the top-right corner.

To assess whether the correlation convergence obtained above follows the theoretical convergence rate, we compare the sampling uncertainty derived from the Fisher-z transform with the quantile approach. Fig. 5.11 presents a theoretical estimate of correlation convergence derived from the Fisher-z transform. The Fisher-z estimate shows s fairly uniform reduction in sampling uncertainty with an increase in both nIBC and nRSD. As the Fisher-z estimate only depends on the input correlation coefficient fields and sample size, combinations with the same total size yield comparable results. For instance, the combinations (2, 6) and (6, 2), as well as (2, 20) and (20, 2), exhibit comparable sampling uncertainty. Thus in this estimate, increases in nIBC and nRSD are comparably effective in increasing sample size.

Comparing the quantile approach (Fig. 5.10) with the Fisher-z approach (Fig. 5.11) reveals that the sampling errors derived from the quantile approach are consistently larger. Therefore, the Fisher-z theory underestimates the sampling error and is not applicable for convective-scale EPSs. In particular, the Fisher-z theory overrates nRSD. In line with the results in Sect. 5.3, increasing nRSD makes copies of samples retrieved by different IBCs and do not contribute to adding information about the distribution of them.

Both approaches emphasise small sampling error in regions with strong correlations. In these highly-correlated regions, the sampling error remains relatively low, around 0.2 for ensemble sizes like (20, 1) and (10, 2), compared to the background correlation coefficient values of approximately 0.7. The current ensemble sizes employed in many operational EPSs (20) appear sufficient to reliably determine the presence of correlations in these highly-correlated regions. In these regions, the last term in Eq. 2.13 is stable, and hence the estimate of total variance becomes reliable if variance sampling uncertainty is small. In contrast, caution is warranted when interpreting regions exhibiting noisy coefficient patterns. In such areas, the last term in Eq. 2.13 significantly varies, hence leading to uncertain total variance. Stated another way, total variance sampling is uncertain if two uncertainty impacts are uncorrelated and additive.

5.5 Optimisation of ensemble design to sample model uncertainty with the smallest ensemble size

While increasing nRSD offers less reduction in sampling error compared to nIBC (as shown in Subsections 5.2-5.4), it effectively captures fine-scale uncertainties not achievable with IBC perturbations. This makes it valuable to find an optimal balance between nIBC and nRSD for limited computational resources.

Start with a simple example. Figure 5.12 shows how the variance of $I_s \times P_t$ changes with the total ensemble size, which is made of a random combination of nIBC and nRSD. For instance, an ensemble size of 10 can be achieved with two combinations of (nIBC, nRSD): (2, 5) or (5, 2). With up to 20 members, the variance of $I_s \times P_t$ grows in parallel with that of I_s but with systematically smaller values (Fig. 5.12a). This suggests that the variance increase is mainly driven by IBCs, and increasing nRSD associated with decreasing nIBC

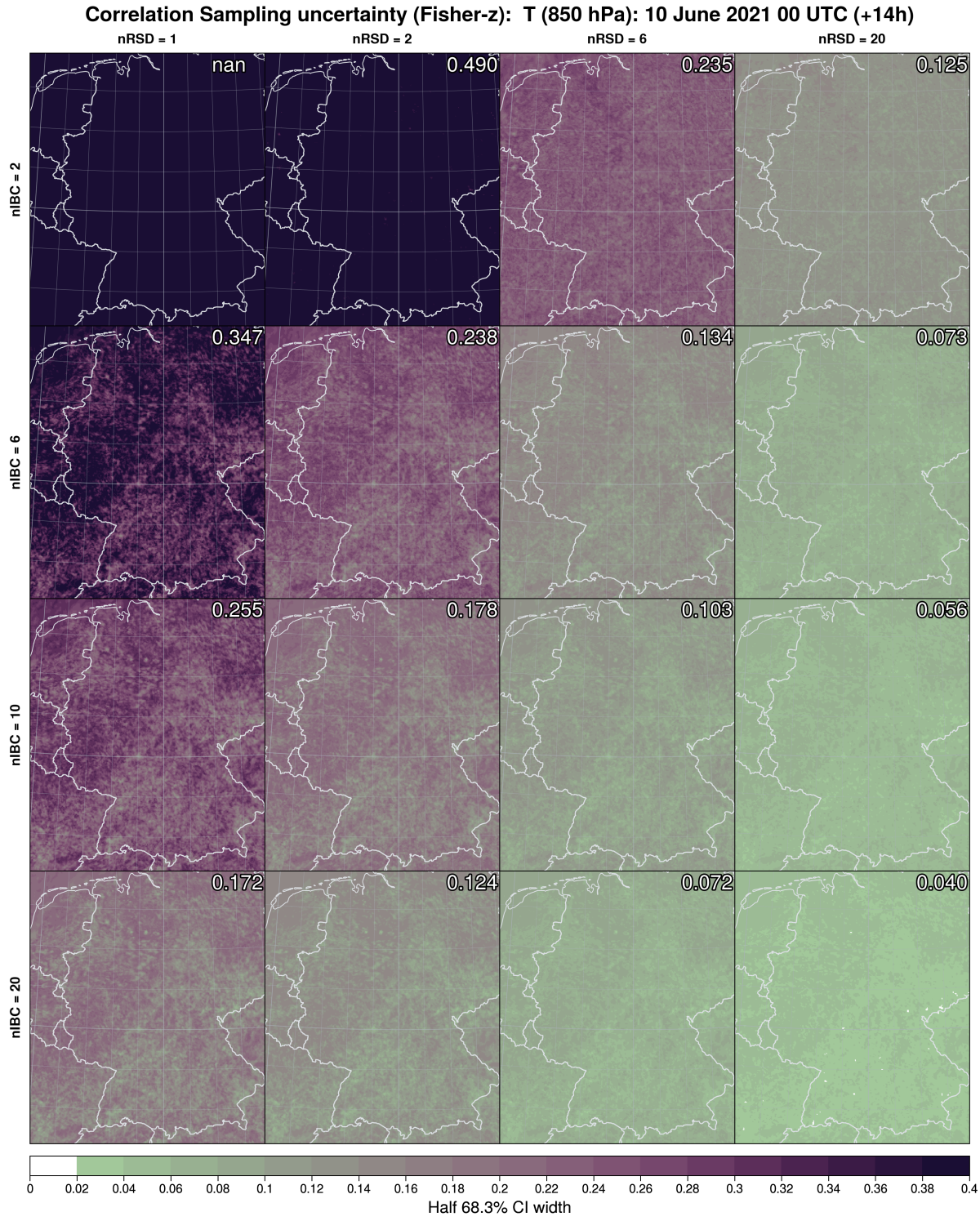

Figure 5.11: As for Fig. 5.10, but for the width of CI estimated using the Fisher-z transform.

systematically reduces the variance; RSD-perturbed members tend to resemble each other. As for sampling uncertainty, the decreasing rates for $I_s \times P_t$ and $P_s|I_t$ are flatter than that of I_s. This is also a result of mixing IBC-perturbed members and RSD-perturbed members equivalently.

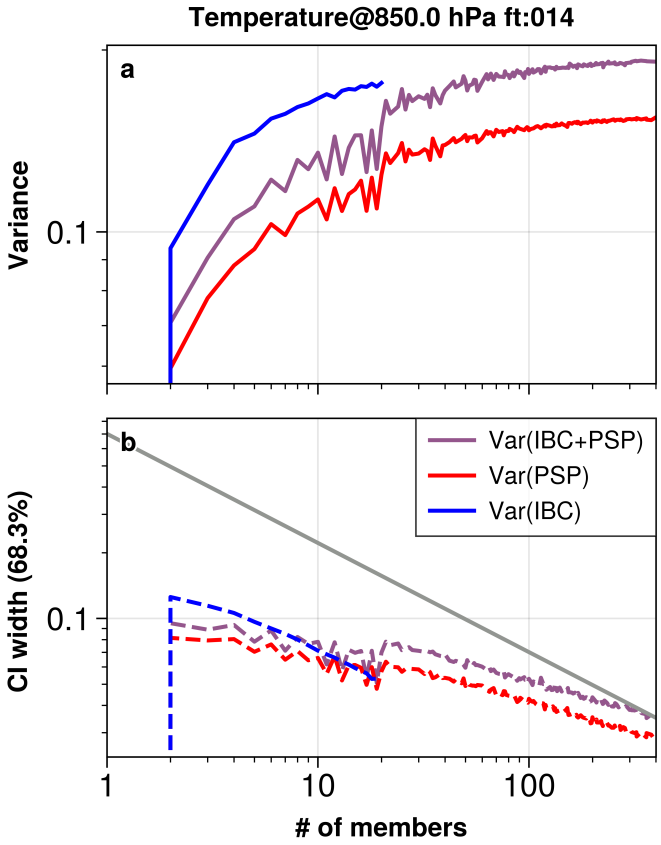

Figure 5.12: As for Fig. 5.4, but as a function of total ensemble size at the time of peak convection (1400 UTC).

A discontinuous change occurs in both the variance and sampling uncertainty at size of 21. From this size onwards, at least two IBCs are always included. This ensures variability in regions where the PSP scheme is inactive, as exemplified by the right-hand side line over northern Germany in Fig. 5.9b. It seems preferable to prioritise a sufficient number of IBC as the foundation of the ensemble design and only increase nRSD when capturing fine-scale variability is crucial.

To identify the optimal combination of (nIBC, nRSD), Figures 5.13 and 5.14 show how variance changes with different combinations of nIBC and nRSD for pre-convection and mid-convection, respectively. Before convection, vertical stripes in $I_s \times P_t$ indicate that only nIBC contributes to the variance (Fig. 5.13a). At this forecast lead time, replacing IBC-perturbed members with RSD ones reduces the ability to capture the full range of possible forecast variability. Once convection occurs, curved stripes in $I_s \times P_t$ with small

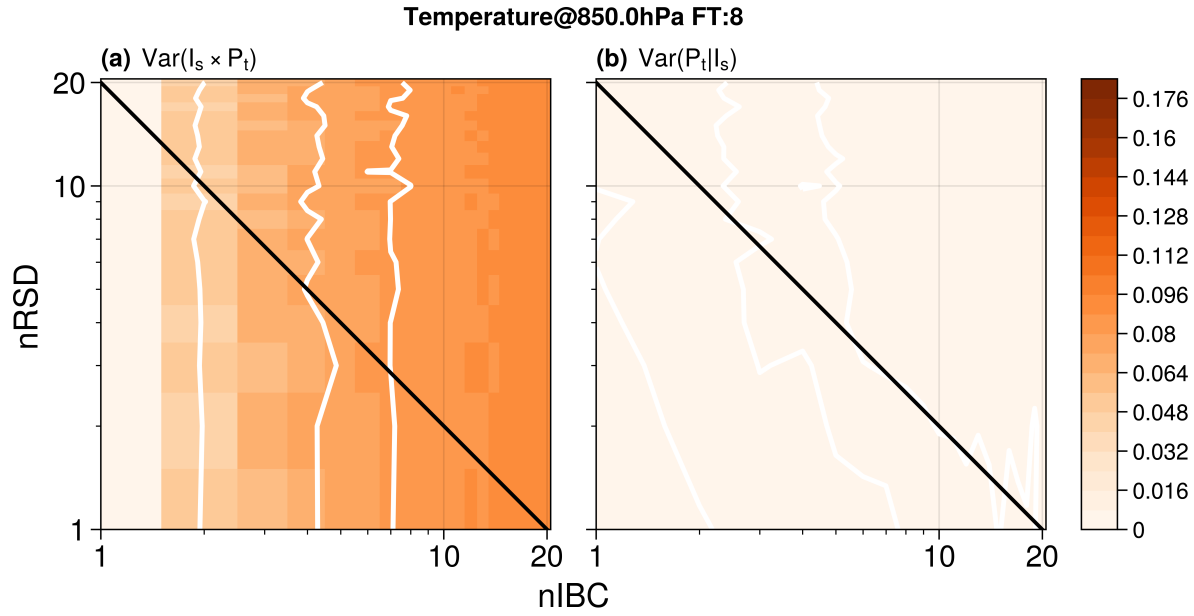

Figure 5.13: 2D variance convergence diagrams before convection (0800 UTC). (a) Total variance and (b) PSP variance are shown in colour. White contours indicate 70 %, 80%, and 90 % isolines of the 400-member variances. Black straight lines indicate combinations of nIBC and nRSD with total ensemble size of 20.

nRSD values (Fig. 5.14a) indicate that nRSD can partially substitute nIBC, up to a nRSD value of 3. There, the contour line of the 90% of the maximum variance parallels the iso-ensemble-size line of 20 members (Fig. 5.14a). As for the variance of $P_s|I_t$, nRSD=2 or 3 shows the largest variance with the ensemble size of 20. (Fig. 5.14b). This suggests potential optimal combinations like (10, 2), (6, 3) or (7, 3) for surely capturing the PSP impact at mid-convection with an ensemble size of 20.

Crucially, while these combinations seem ideal for mid-convection, they might negatively impact variance when convection is absent, highlighting a significant trade-off. In addition, the increase in nRSD is not beneficial for more accurate sampling of strong correlation. Finding the optimal balance between nIBC and nRSD necessitates accounting for the forecast stage and its associated uncertainties. Considering that convection is not active for most of time, optimal for long term assessment would be to keep nIBC as high as possible with one RSD each.

5.6 Summary and Conclusions

This chapter serves to to add knowledge about convergence of ensemble variability when a new model uncertainty source is introduced. We investigates sampling uncertainty of variance of model uncertainty using the variability budget method, which decomposes total variance into sum of individual variances and their interactions. To assess the influence of

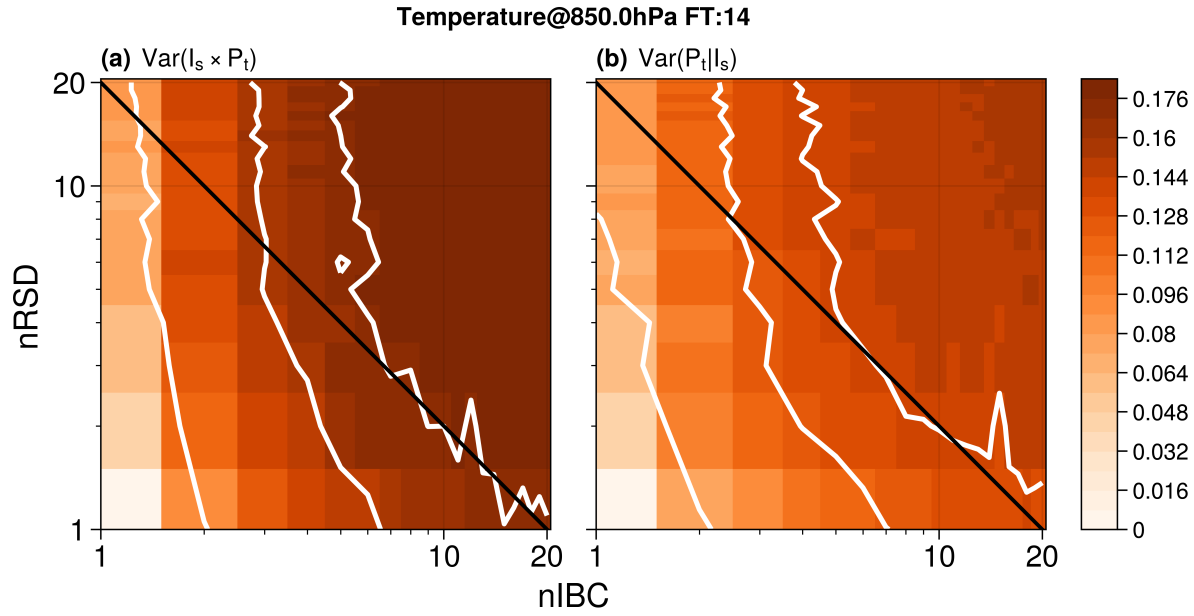

Figure 5.14: As for Fig. 5.13, but at the peak convection (1400 UTC).

ensemble size given by different perturbation methods, a large ensemble simulation with 400 members, a combination of 20 IBCs and 20 RSDs, is performed. Smaller ensembles are made by bootstrapping (repeating sub-sampling with replacement) the large ensemble. First, we estimate the necessary nIBC and nRSD to ensure the sampled variance closely approximates true variability, in comparison with the scaling theory shown in Craig et al. (2022) and Tempest (2023). Second, we investigate convergence of correlation coefficients between IBC and PSP impacts. To find out whether sampling uncertainty of correlation scales with an asymptotic theory as shown for variance, the direct sampling results by bootstrapping are compared with the theoretical estimates based on the Fisher-z transformation. Finally, we suggest optimal combinations of ensemble sizes for nIBC and nRSD for effective variability sampling.

We begin by exploring the evolution of total variance as a function of nIBC and nRSD. The necessary ensemble size for accurate variance estimation depends on variables under consideration and the presence of convective activity. Temperature variance above the boundary layer converges more rapidly, requiring around nIBC=6, while variables like U-wind necessitate a larger ensemble size of approximately nIBC=10 to achieve saturation. Both variance convergence exhibit a steady decrease in sampling uncertainty proportional to $nIBC^{-1/2}$, consistent with findings for convective-scale ensembles by Tempest (2023). This difference can be attributed to the smoothing effects of atmospheric dynamics on temperature fields, which lead to faster convergence compared to wind variables.

The impact of increasing RSD becomes evident only after the onset of convection, with nRSD of 2-3 being sufficient to capture variance attributed to the PSP scheme. The effectiveness of increasing nRSD is limited compared to nIBC, as the latter can encompass

the range of PSP uncertainty with a sufficiently large nIBC (> 10) used. Nevertheless, increasing randomness in the PSP scheme introduces fine-scale variability that cannot be achieved through IBC perturbations alone, particularly in regions dominated by local convective activity. Increasing nRSD seems to selectively enhance convective-scale variances, but it is inactive elsewhere and does not increase sample size as effectively as nIBC does.

The convergence of correlation estimates between IBC and PSP impact is significantly influenced by the underlying correlation strength. Regions where the PSP scheme is active exhibit strong negative correlations and faster reduction in sampling uncertainty. In contrast, areas with weak or noisy correlations require larger ensemble sizes to stabilise the correlation estimates. Comparing sampling uncertainty obtained through bootstrapping with the theoretical estimates based on the Fisher-z transform reveals that the convergence of correlation in practice is slower than the theoretical expectations. Like the convergence of the extreme percentile values in Craig et al. (2022), we need larger nIBC than the theoretical estimate to accurately sample the interaction between the IBC and PSP impacts. Nevertheless, the benefit of nRSD in reducing sampling uncertainty is limited compared to nIBC, like the finding for the variance convergence.

Finally, we discuss optimal combinations of nIBC and nRSD for practical forecasts, in which total ensemble size is constrained by computational resources. Having just one RSD already captures 90% of the maximum total variance, and nRSD = 2 and 3 are optimal for sampling variability due to the PSP scheme once convection is active. However, trading off nIBC for nRSD significantly compromises variability sampling in absence of convection. Therefore, for long-term assessments, prioritising higher nIBC is recommended unless computational resources allow for substantial increase in the total ensemble size.

This study mainly investigates how standard deviations converge, which corresponds to the 68.13% width of the distributions, assuming they are perfect Gaussian distributions centred at the ensemble mean. By multiplying Eq. 2.13 by a factor of 4 and 9, one can easily determine the widths for the central 95.4% or 99.7% of the distributions. Thus, this method appears to be easily extended to assessing a variability budget for "extreme" weathers, which are often defined as 95%- or 99%-ile values. In such a case, the relative contributions of individual factors remain unchanged as obvious from Eq. 2.13. Although this application sounds very useful and straightforward, nevertheless, extra caution is warranted when examining the convergence of these extreme values. These extreme values are generally more affected by non-linearity of the governing equations. In highly non-linear weather models, these extreme percentiles converge much more slowly than the standard deviation (Tempest et al., 2023). To properly analyse the convergence of such extreme quantiles, a bespoke convergence study with much larger ensemble size is necessary.

It's important to note that these results are based on a single case study under weak convective forcing condition. The necessary size of nIBC may vary in different forcing regimes and seasons. However, the minor importance of nRSD to nIBC is likely to remain small. Model uncertainty representations are able to suggest another scenario in the given atmospheric states, but are NOT able to create, for example, a thunderstorm in a region characterised by the stable troposphere. This limitation applies, at least to some extent, to other model uncertainty schemes as well. As long as model uncertainty is acting in

sub-processes of the activity constrained by IBC, a lower level of randomness in a model uncertainty scheme is expected to be sufficient.

Chapter 6

Quantifying uncertainties with the variability budget

The previous chapters addressed the model uncertainty impact on increasing country-scale variability (Chapter 3) and scale-dependent variability (Chapter 4) for convective-scale forecasts. This chapter aims to trace those impacts to their origin by tracking evolution of the variances of the impacts and their interactions, addressing the last research question **How does forecast uncertainty from different sources evolve and interact over time?** The results from Chapter 5 support to structure an ensemble dataset for efficient sampling of model uncertainties, as well as to validate robustness of sampled correlations.

The dataset is described in Sect. 6.1. In Sect 6.2 the time evolution of variability and interactions is presented in a flow-dependent manner. The features of model uncertainties leading to interactions are assessed in Sect. 6.3. The effect of the order of adding uncertainty representations is examined in Sect. 6.4. The findings are summarised in Sect. 6.5.

6.1 Regional grand ensemble

6.1.1 Simulation design

For a comprehensive investigation of the flow-dependent impact of IBC, PSP, and MPP, we conducted one-month routine simulations of an "ensemble of ensembles," termed the regional grand ensemble (RGE). The RGE was basically designed as an enhancement of the trial run dataset (Sect 4.3.1). It consists of daily ensemble simulations initialised every 00 UTC from 15 August to 14 September 2023, with a 24-hour forecast lead time. The ensemble design employs the cycle-equitable graph with the simplified Morris concept, as described in Section 2.5. Each day's RGE comprises four 20-member ensembles, incrementally adding model uncertainty representations. The first ensemble, denoted I (referencing IBC), includes only IBC perturbations and serves as the baseline. In the IP and IM ensemble, the PSP scheme and parameter perturbations in microphysical parameters are turned on, respectively. The combined ensemble IPM incorporates all three uncertainty

representations.

Perturbations in IBC and of PSP follow the approach used in the trial run (Sect 4.3.1). ICs are provided by the operational ensemble analysis of the ICON-D2-KENDA system. LBCs are from the operational ICON-EU-EPS output. The first 20 members of them were used for the initialisation. When the PSP scheme in turned on (in the IP and IPM ensemble), it introduces spatial variability through a random seed mechanism. This seed is assigned sequentially, changing from 1 to 20 across each ensemble member.

To efficiently capture the microphysical impact, we simplified MPP from the case studies used in Chapter 3 and 4. This simplification involved constructing 20 unique combinations of CCN concentration and CDSD shape parameters. We selected four distinct CCN concentrations and five CDSD parameters to generate this diversity. It's important to note that the default CCN configuration required adjustment due to an update in the ICON model. The new default is $N_{CN} = 250cm^3$, thus our chosen CCN values include $N_{CN} = 100cm^3$, $250cm^3$, $1700cm^3$, and $3200cm^3$. The CDSD shape parameters are chosen from 0, 1, 2, 4, and 8. When the MPP scheme is activated in the IM and IPM ensemble, each of these 20 combinations is assigned to one ensemble member. While a specific combination remains fixed for a particular IBC for simplicity, this choice doesn't impact overall ensemble variability because of inherent randomness and interchangeability of IBC in the setup of the ICON-D2-KENDA.

6.1.2 Classification of forcing regimes

Throughout the period, all days received precipitation somewhere within the German domain, although for many days the precipitation occurred in trace amounts or only in small regions. After excluding days with less than 1 mm/d accumulation on average, we classified convective regimes for 19 days using the convective adjustment timescale (τ_c; Sect. 2.3). Two cases that showed a clear diurnal cycle of precipitation and a clear τ_c peak before daytime convection were classified as weak forcing cases (red dots in Fig. 6.1). Days with continuous organised rainfall and daily mean τ_c less than 0.6 hours were classified as strong forcing conditions, resulting in 7 days (blue dots in Fig. 6.1).

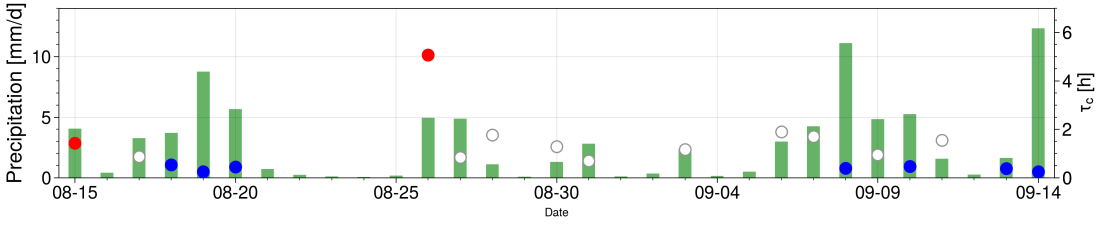

Figure 6.1: Time series of the day-to-day variability of 24-h accumulated precipitation (bars) and convective adjustment timescale τ_c (dots). The colours of the dots represent weak (red), intermediate (white) and strong (blue) forcing regimes. Green bars depict I ensemble mean area-averaged rainfall.

We first selected a typical case for each forcing regime to showcase detailed results. The 15th of August represents a weakly forced case, characterised by a subtle horizontal equivalent potential temperature gradient (Fig. 6.2a) and weak westerly winds flowing from eastern France towards Germany (not shown). The timeseries of precipitation rate followed a clear diurnal cycle, while the spatial distribution of accumulated precipitation exhibited intermittency, particularly in northern Germany (Fig. 6.2c). The 19th of August represents a strongly forced case, featuring a strong, zonal gradient of equivalent potential temperature over the evaluation domain especially along the edge of the precipitated region (Fig. 6.2b). Relatively cold air advected from the northwest led to a temperature inversion in southern and eastern Germany, fostering widespread, organised precipitation that began in the morning and persisted throughout the day (Fig. 6.2d). We will initially focus on the results at 1600 UTC for these two days to exemplify flow-dependent behaviours after substantial precipitation. An aggregated analysis over all seven cases within each regime is presented in Section 6.2.

6.1.3 Application of the variability budget to RGE

To assess the evolution of variability and interactions between uncertainties from the three sources of uncertainties, IBC, PSP and MPP, we compute each term in Eq. 2.14 with $n = 3$:

$$\begin{aligned} Var(x'_{IPM}) = {} & Var(x'_I) + Var(x'_P) + Var(x'_M) \\ & + 2[Cov(x'_I, x'_P) + Cov(x'_I, x'_M) + Cov(x'_P, x'_M)], \end{aligned} \tag{6.1}$$

where x'_{IPM} represents total impact defined as a deviation of an ensemble member in the IPM ensemble from the ensemble mean, corresponding to debiased mixed effect EE_{IAB} in Eq. 2.10. x'_I, x'_P, and x'_M are debiased impacts of IBC, PSP and MPP, respectively, corresponding to debiased EEs in Eq. 2.9. These debiased perturbations are defined as:

$$\begin{aligned} x_i'^I &= f_i(0,0) - \overline{f_i(0,0)} \\ x_i'^P &= f_i(1,0) - f_i(0,0) - \overline{f_i(1,0) - f_i(0,0)} \\ x_i'^M &= f_i(1,1) - f_i(1,0) - \overline{f_i(1,1) - f_i(1,0)} \end{aligned} \tag{6.2}$$

where $\overline{\ }$ indicates average, the subscript i refers to the ith ensemble member and $f_i(PSP, MPP)$ is a model output value derived with the three input factors perturbed in RGE. The PSP and MPP flags here are either 0 (turned off) or 1 (turned on). In the original Morris method, x_i^M should be normalised by difference in the parameter values (Δ in Eq. 2.8) to approximate local partial derivative. However, we skip this step to focus on the impact given by absence or presence of uncertainty. In practice debiasing impact has no effect on the calculation of variances in Eq.6.1, but we keep subtracting the ensemble means of x^P and x^M in the second and third lines in order to have consistency with the definition of x'_* in Sect. 2.5.2 and Figure 2.4.

In practice, x'_* are calculated by getting a change in deviation from the ensemble mean

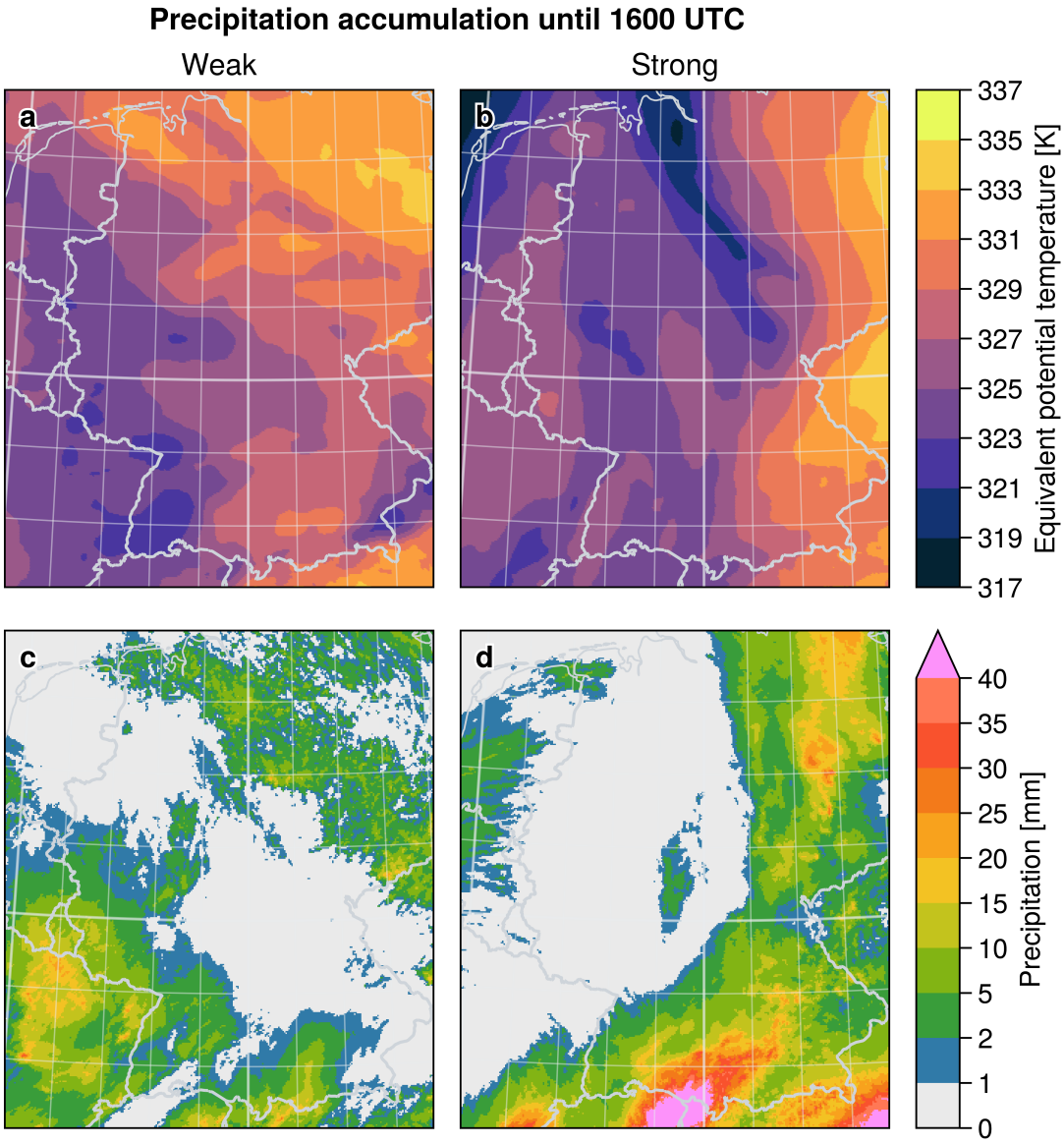

Figure 6.2: (a, b) Equivalent potential temperature at 850 hPa height and (c, d) Accumulated precipitation until 1600 UTC for (a, c) 15 August 2022 and (b, d) 19 August 2022. The values are the ensemble mean of the I ensemble.

by adding uncertainty representations:

$$\begin{aligned} x_i'^{I} &= I_i - \overline{I} \\ x_i'^{IP} &= IP_i - I_i - \overline{[IP - I]} \\ x_i'^{IM} &= IPM_i - IP_i - \overline{[IPM - IP]} \\ x_i'^{IPM} &= IPM_i - \overline{[IPM]}. \end{aligned} \tag{6.3}$$

The variance terms in Eq. 6.1 can be easily obtained, and the covariance terms can be calculated directly or indirectly using the relationship in Eq. 2.15 to save computational resources.

The formulation in Eq. 6.3 defines the elementary effect of MPP with the IP ensemble as a starting point, while the PSP effect is defined with the starting point of I. This order is irrelevant if the elementary effect is independent of the starting point. However, in nonlinear models like NWP models, this dependence often occurs. To assess this effect of the starting point, we calculate the terms in Eq. 6.1 along a different pathway in the cycle-equitable graph (Fig. 2.3b). We refer to the original pathway used in Eq. 6.3 as the "P-path" and the alternative pathway as the "M-path," where the effects $x_i'^{P}$ and $x_M'^{I}$ are calculated as:

$$\begin{aligned} x_i'^{P} &= IPM_i - IM_i - \overline{[IPM - IM]} \\ x_i'^{M} &= IM_i - I_i - \overline{[IM - I]}. \end{aligned} \tag{6.4}$$

The results presented in the following subsections primarily utilise the P-path for display and discussion except for Subsect. 6.4 in which we compare the M-path results with the P-path. Additionally, for brevity, we use the first letter of each uncertainty representation to denote its impact. For instance, $Var(I)$ means $Var(x'^{I})$.

6.2 Flow-dependent evolution of impact

This subsection investigates how individual impact variances evolve, interact, and contribute to the increase in total variance. First we start with assessing a net variance increase obtained by adding model uncertainty representations to verify whether there is correlation between impacts of different sources. If the impact of model uncertainty is independent from the IBC uncertainty impact, they are being uncorrelated, and hence their respective variances will be linearly added, as presented in Sect. 2.5.

Figure 6.3 shows time series of ensemble variances for respective ensembles in the RGE on the representative cases for respective forcing regimes. Overall, the net variance increase is larger during weak forcing and reaches a maximum of around 15 % around noon for specific humidity in weak forcing. This value is considerably smaller than the expected linear sum of respective variances, as estimated, for example, using the work by Keil et al. (2019) (a ratio of squared normalised spread in Fig.5), where the individual variance due to PSP and microphysics (CCN) is about 70 % and 20 % of the EPS variance, respectively.

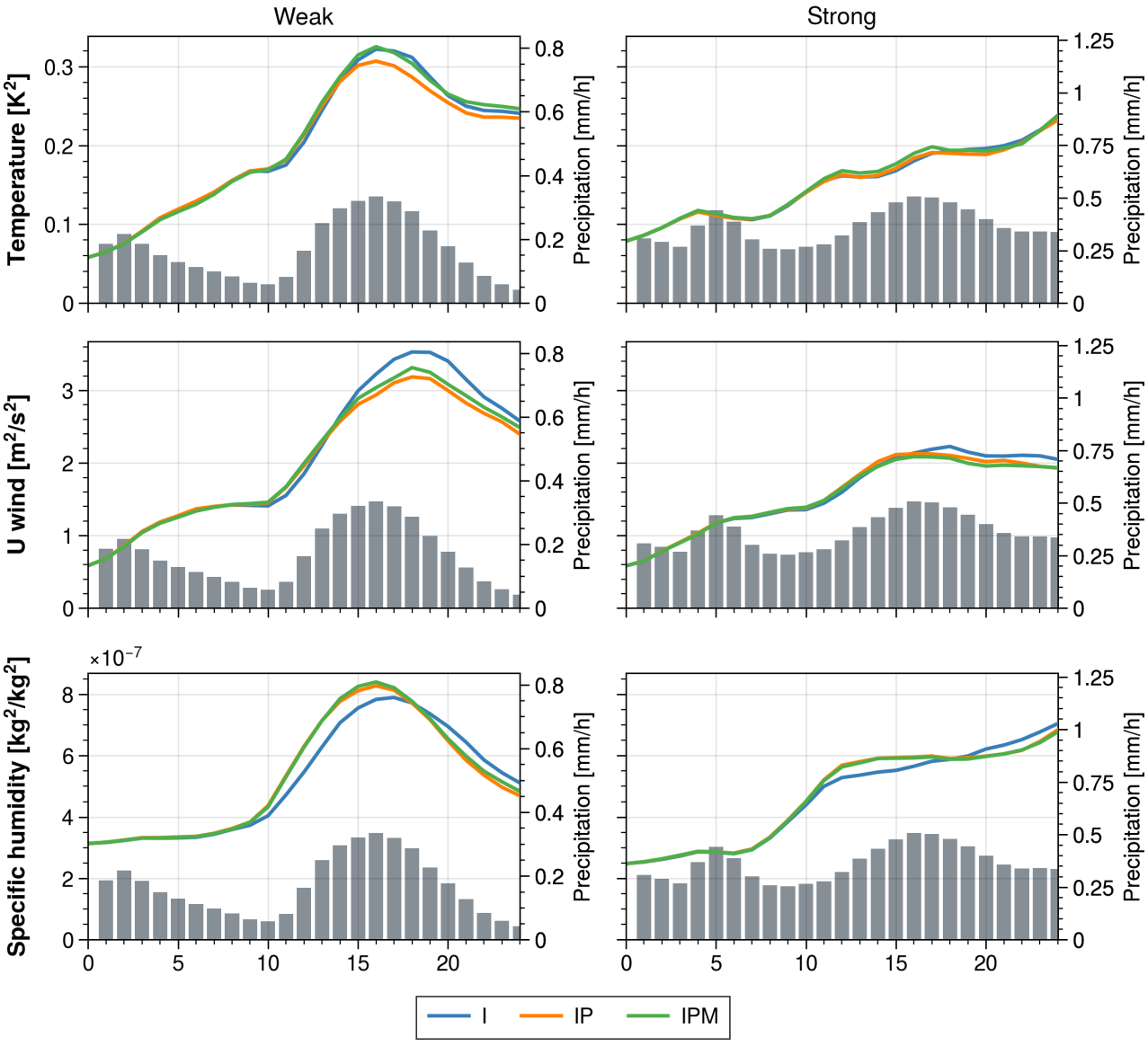

Figure 6.3: Time series of total variance for the ensembles (blue) I, (orange) IP, (blue) IM, and (black) IPM. Variance at 850 hPa pressure levels for the weak forcing (left, 15 August) and strong forcing (right, 19 August) case are shown. Bars show hourly precipitation amounts over the evaluation domain, showing the diurnal cycle of convection.

The discrepancy between the obtained and expected variance increase suggests the presence of correlations between uncertainty impact, which contributes to suppressing total variance increase.

Figures 6.4 and 6.5 present individual contributions of each term in Eq. 6.1, showing that the contributions of uncertainty sources (IBC, PSP, MPP) evolve differently across variables, heights, and convective regimes. For temperature and wind in the mid and lower troposphere, both PSP and MPP impacts increase from 0600 UTC, growing steadily before rapidly rising in the afternoon. In the upper troposphere (200 hPa), their contributions are invisible until 1000 UTC but then increase rapidly. Initiation of deep convection facilitates the upward propagation of PSP and MPP impacts from lower levels. Once upper and lower layers are linked by convection, uncertainty grows at a similar rate in both forcing regimes. However, the relative importance of IBC uncertainty is significantly larger in strong forcing, where large-scale forcing is perturbed by perturbations in IBC. This finding aligns with existing studies on model uncertainty growth in different convective regimes (Flack et al., 2018; Keil et al., 2019; Johnson and Wang, 2020).

In both forcing regimes, adding model uncertainty representations only marginally increases the total variance compared to summing individual variances. The covariance terms $Cov(I, P)$ and $Cov(P, M)$ evolve with the model uncertainty variances and almost cancel out their impact. Interestingly, while the variances decrease after convection, the covariance terms diminish less significantly once saturated. For instance, $Var(I)+Var(P)$ of 850-hPa temperature in weak forcing drops by 0.1 K^2 from 1600 UTC to 2400 UTC, but $Cov(I, P)$ only drops by 0.02. This leads to the net variance decrease in Fig. 6.3. The implication is that correlation has a saturation level and it remains saturated once reached. If the model uncertainty impact can grow rapidly and continue to grow after correlation saturates, it could increase total variance efficiently.

Model uncertainty variances for specific humidity, on the other hand, start growing in the lowest layer and grow faster at higher layers during weak forcing (Fig. 6.4, rightmost column). The PSP scheme produces slight gains in the total variance across layers during the rapid growth around noon, followed by an increase in the covariance term that cancels out the impact. This different growth behaviour is physically reasonable since fast-propagating modes like gravity waves will quickly lift disturbances in the planetary boundary layer and cause simultaneous variance and covariance rises in temperature and wind across layers, while changes in moisture variables require advective transport.

Although the evolution the individual variances varies slightly by level and variable, their common behaviour can be summarised as follows. Model uncertainty for convective-scale forecasts starts acting with initiation of convection in the lower troposphere. The impact propagates quickly upwards for dynamics variables, while requiring advective transport for moisture variables. Although the impact on total variance may be invisible or small, this does not mean model uncertainty representations are inactive, but rather that their effect is cancelled by strong correlation. Once correlation reaches saturation, it remains saturated after convection ceases. One notable benefit of the PSP scheme, however, is that the scheme gains the total variance during convection initiation, meanwhile MPP prolongs the lifetime of variance in weak forcing regimes, in which time ensemble forecasts

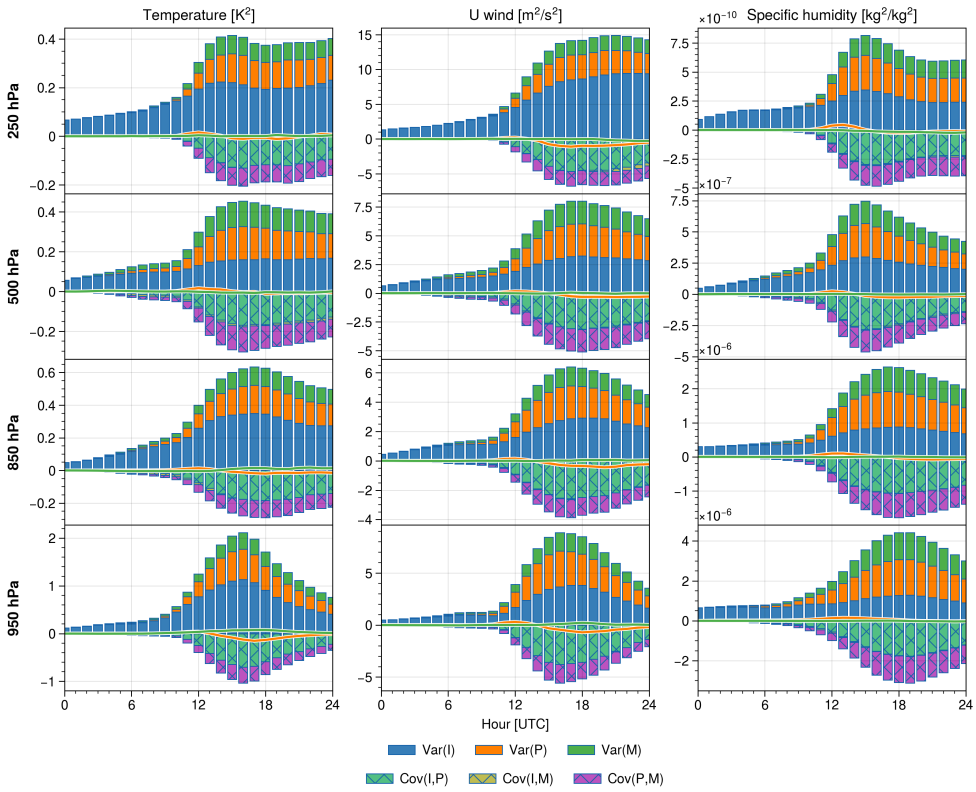

Figure 6.4: Temporal evolution of variance terms and covariance terms in Eq. 6.1 averaged over two weak forcing cases for (left column) temperature, (middle) u component of wind and (right) specific humidity. Values are averaged over the German domain. Orange and green lines indicate the differences between the IP and I, and IPM and IP in Fig. 6.3, respectively, indicating small net impact on the total variances compared to the individual variances.

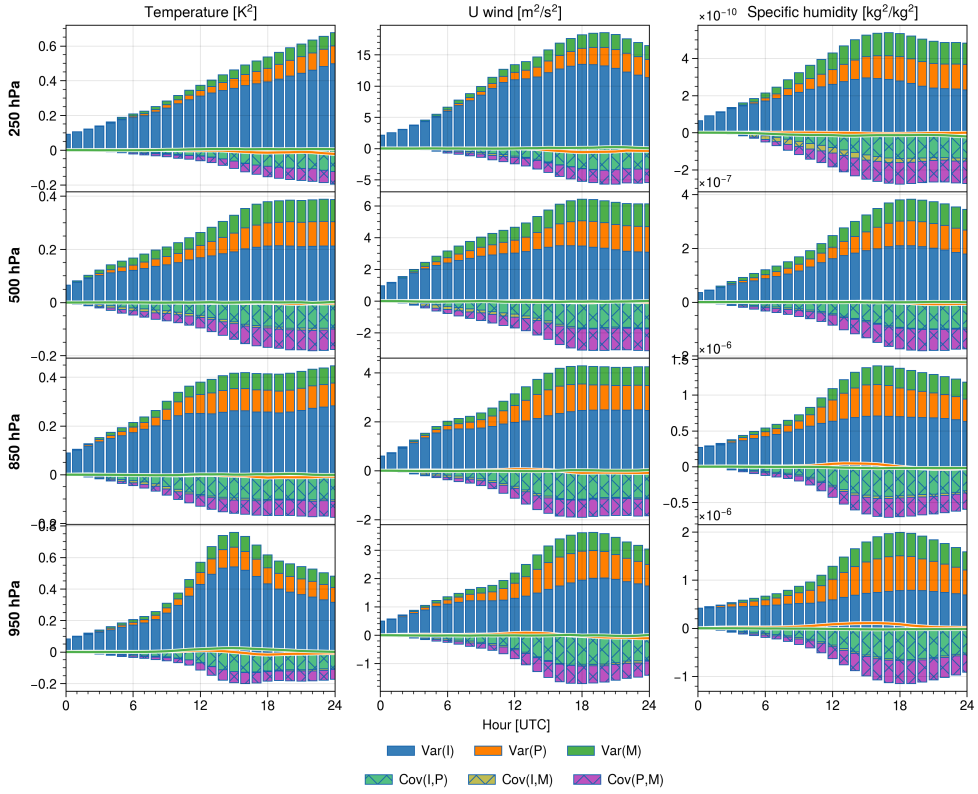

Figure 6.5: As for Fig. 6.4, but for seven strong forcing cases.

are known to be underdispersive. This emphasises model uncertainty representation's ability to focus on improving representation of bespoke events (Plant et al., 2022).

6.3 Structure of variance and correlation

To understand how variance is cancelled out, this section further investigates the structure of variance and correlation in physical and probability space by focusing on representative cases for the two convective regimes. We look at temperature at 850 hPa level since it is closely related to warm uplift initiating convection as well as found to be sensitive to model uncertainty (Romine et al., 2014). Since the total variance is unchanged until 1500 UTC and after that the model uncertainty impact starts to be visible in Fig. 5.2, examining behaviours at 1600 UTC seems suitable, as the presence of the impact and correlation is certain but cancellation is strong at that time. Interestingly, this time coincides with the peak time of hourly precipitation for both cases.

Figure 6.6 contrasts the spatial distribution of variances for the weak forcing case. The IBC variance exhibits a wide spatial extent across the domain, with strong signals concentrated in three northwest-southeast bands (Fig. 6.6a). The northern and southern bands, located between Hamburg-Berlin and along the Alps, respectively, co-locate with

regions exceeding 1 mm/d accumulated precipitation (Fig. 6.6d). The central band, which co-locates with a weak equivalent potential temperature gradient in Fig. 6.2a, lacks any significant precipitation association. The PSP variance exhibits strong signals along the northern and southern bands, but not in the central band, suggesting its selective activation in regions of convective activity (Fig. 6.6b). The PSP scheme's selective activation is attributed to its ability to initiate convection in unstable regions. While also showing selective behaviour, the MPP variance displays a weaker and more scattered pattern, (Fig. 6.6c) likely due to its primary effect on altering the amplitude of convection rather than activating convection (discussed in Chapter 4). The model uncertainty representations generally become active within subregions of the IBC impact, in line with the results in Frogner et al. (2022).

Under strong forcing, IBC perturbations have pronounced impact over eastern Germany, coinciding with regions of precipitation (Fig. 6.7a). The MPP variance closely aligns with the precipitation over eastern Germany (Fig. 6.7c, d), while the PSP variance shows a more uniform distribution across Germany with modest amplitudes (Fig. 6.7b). Both variances are smaller compared to the weak forcing case, and the PSP impact appears to arise from a random growth of small perturbations rather than the selective effect found for the weak forcing case. This small model uncertainty impact can be explained by the strong synoptic modulation dominated by IBC, which largely controls convection leaving little room for model uncertainty to further initiate convection. As with weak forcing conditions, model uncertainty representations are active within regions where the IBC impact is active during strong forcing, but not as systematically as weak forcing conditions.

To summarise the spatial distribution of variances, the variances of PSP and MPP are consistently active in regions where the IBC variance is present, suggesting that in regions where model uncertainty representations are active, the IBC uncertainty is always also active (confirmed across various lead times, not shown). The evolution of variance is linked to unpredictable weather events rather being driven by a specific parameterisation scheme. If the background atmospheric state, acting as an input parameter for a scheme, is stable and predictable (i.e. showing no IBC variance), the model uncertainty does not act in that region.

Next, we investigate the spatial distributions of correlations between the three uncertainty impacts (Fig. 6.8 and 6.9). Grid points with less than 0.1 K^2 variances for any of the three uncertainties are excluded for clarity. The correlation between IBC and PSP exhibits a systematic negative pattern, particularly pronounced in the northern band and over the Alps (Fig. 6.8a), where the PSP variance is largest and local convection dominates. Regions with smaller PSP variance, like the border between France and Germany, display a mix of negative and positive correlations without a clear pattern. The average correlation coefficient is around -0.45. This negative correlation implies that when an ensemble member in the I ensemble has a higher temperature than the ensemble mean, adding the PSP perturbations brings it closer to the mean in the IP ensemble, and vice versa. Note that average correlation coefficients should generally be calculated from aggregated covariances divided by root-mean-squared difference (Déqué, 2011), a simple spatial average of correlation coefficients is used here for straightforward interpretation of point-wise

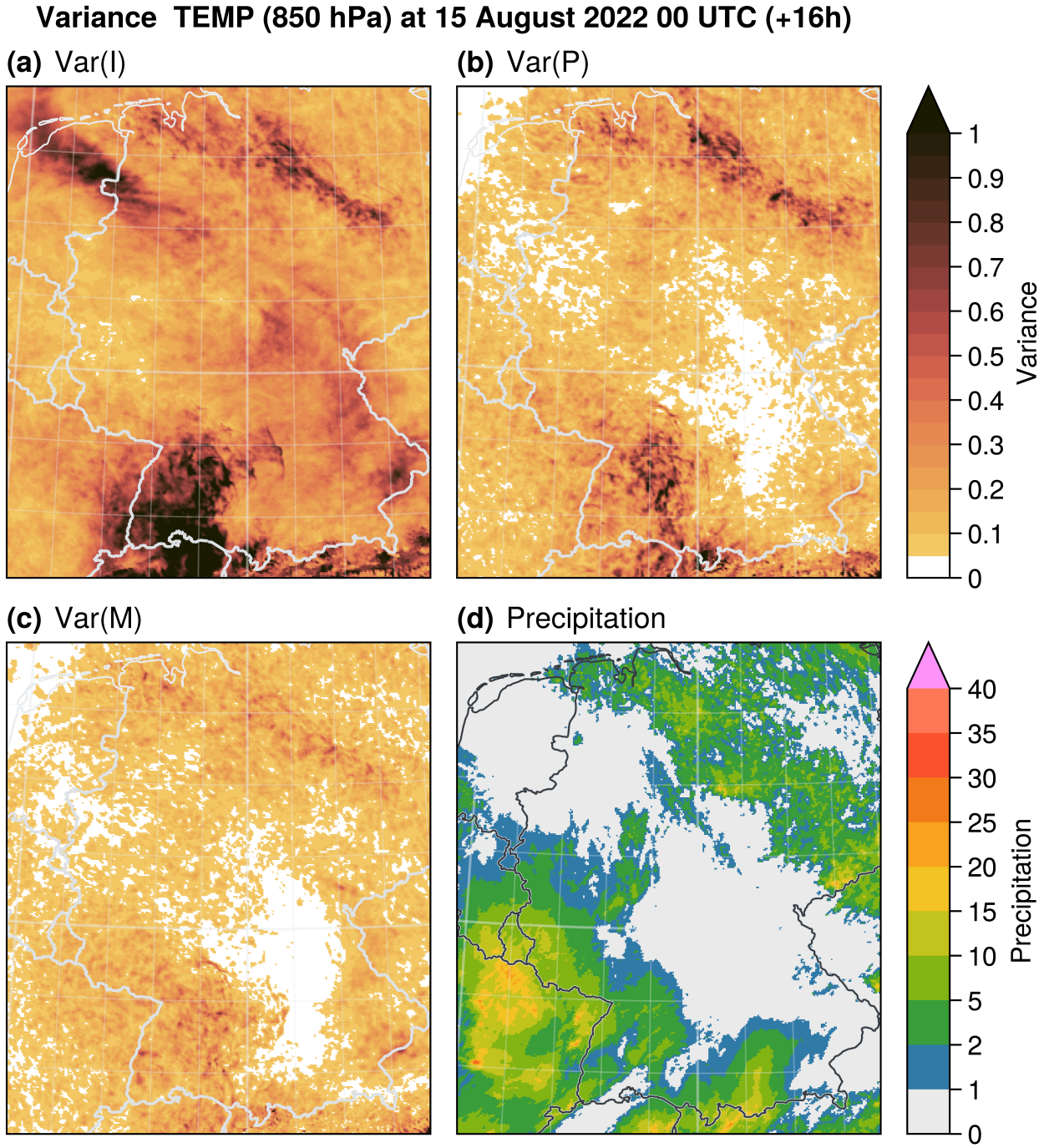

Figure 6.6: Variance maps of the (a) IBC, (b) PSP and (c) MPP impact for weak forcing (1600 UTC on 15 August). (d) Precipitation accumulation, same as Fig. 6.2c for ease of comparison.

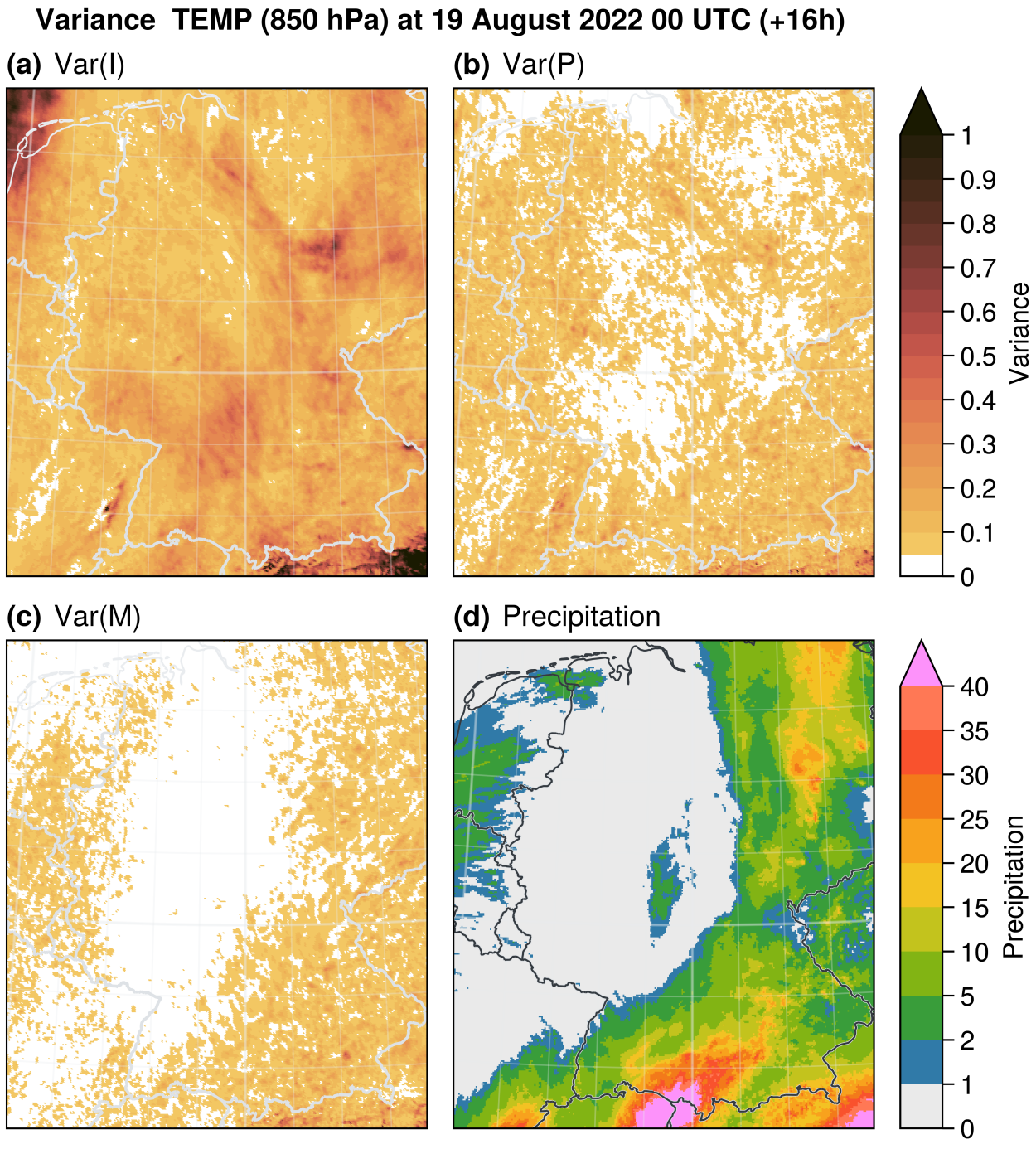

Figure 6.7: As for Fig. 6.6, but for strong forcing (1600 UTC on 19 August).

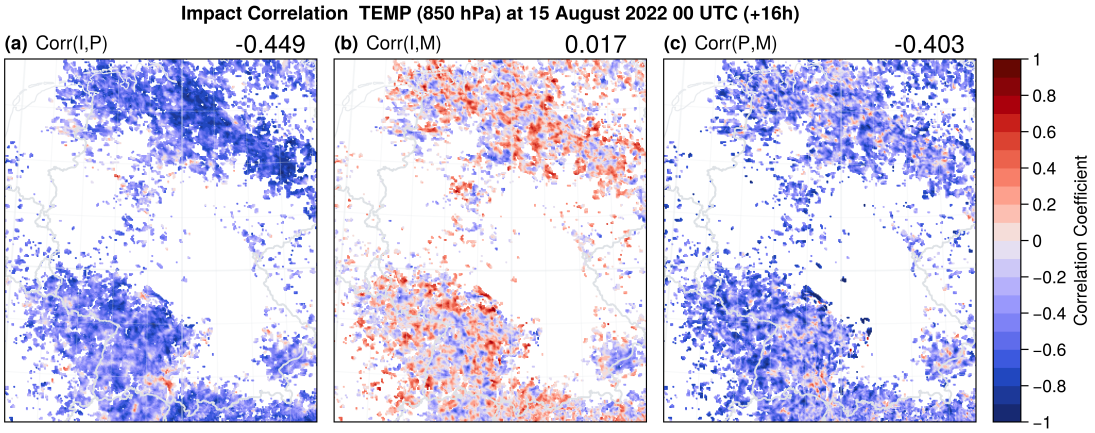

Figure 6.8: Correlation coefficient maps of IBC-PSP (top), IBC-MPP (middle) and PSP-MPP (bottom) calculated within the P-path for the weak forcing case (1600 UTC on 15 August). The area-averaged correlation coefficient values are shown at the top right corner. Grid points where either three variances are less than 0.1 K^2 are masked.

correlation.

In contrast, the correlation between IBC and MPP displays a noisy pattern of strong correlations and no correlation when averaged across Germany (Fig. 6.8b). The spatial scale of the signals matches the scale of the MPP variance shown in Fig. 6.6c (also the scale of locally triggered convection) and the scale of typical displacement of precipitation cells discussed in Sect. 4.2.4. This suggests that the IBC-MPP correlations at grid-point scales are related to response of individual convective cells, but the response is fairly random.

The correlation between the model uncertainty impacts (PSP and MPP) is systematically negative, but exhibits a noisier pattern compared to the IBC-PSP correlation (Fig. 6.8a and c), suggesting smaller effective scales for model uncertainty impacts. If we think of a simple combination of mean correlations of IBC-PSP and PSP-MPP, it suggests that where IBC perturbations leads to higher temperatures than the mean, PSP perturbations tend to cool them down, and MPP might heat them again. This sounds as if it implies a positive correlation between MPP and IBC, but this is not the case despite the fact that these correlations are found on the same grid points (as shown in Fig 6.8b). What perturbations the IBC uncertainty add does not systematically change the direction of the MPP impact.

Despite differences in how variances evolve between weak and strong forcing regimes, the respective correlation behaviours during strong forcing remain consistent, showing weaker IBC-PSP correlation and stronger PSP-MPP correlation (Fig. 6.9a-c). Therefore, the factors determining correlations appears to remain unchanged between forcing conditions.

The investigation of correlation maps already inferred the existence of linear relationships between the impacts. To further elucidate these structures of correlation, Fig. 6.10 shows cross correlations of the impacts aggregated over all grid points where all three variances are larger than 0.1. First of all, the diagonal panels (Fig. 6.10I, II, III) show that

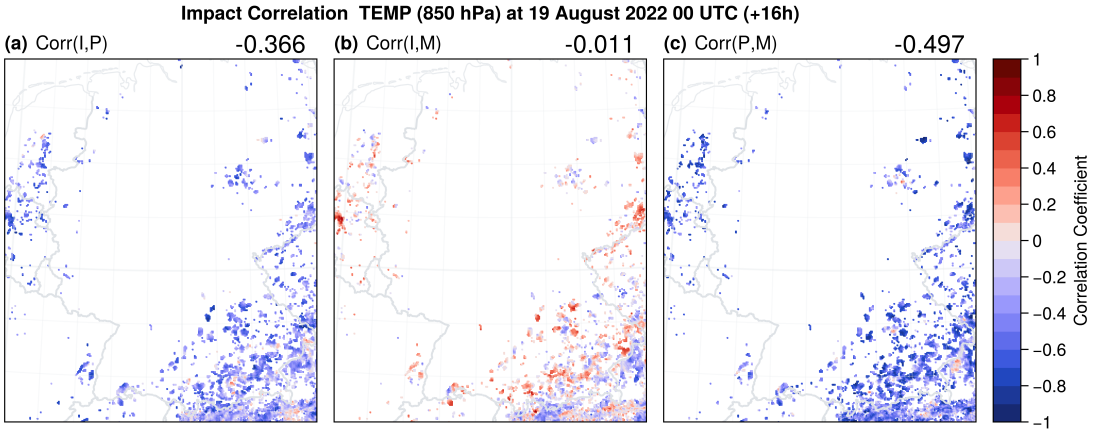

Figure 6.9: As for Fig. 6.8, but for strong forcing (1600 UTC on 19 August).

individual impacts follow Gaussian distributions with zero mean (this is a result of debiasing in Eq. 2.9) when aggregated over the domain. However, the distributions become skewed in response to the correlated impact as discussed later.

The IBC-PSP scatter density shows the linear relationship of their impacts, which means when IBC impact is positive and large, PSP impact is likely to be negative and large (Fig. 6.10a). However, on average, only one-thirds of the IBC impact is cancelled out by this linear relationship, as the linear regression line shows the slope of -36.3 %. This linear relationship is also found for the PSP-MPP correlation (Fig. 6.10c) with correlation coefficient of -0.373 and the regression slope of -30.3%, as well as for upper layers with stronger correlation, such as -0.571 for IBC-PSP correlation at 500 hPa level (not shown). However, the linear relationship is not evident in the correlation between IBC and MPP (Fig. 6.10b). While the sign of MPP impact is independently (randomly) determined from the sign of IBC impact, the amplitude of MPP impact is strongly constrained by IBC; MPP shows systematically small impact as IBC shows larger. This contrasting response raises two sub-questions: 1) from which does this linear relationship arise? and 2) from what does the remaining two-thirds arise?

The linear relationship appears to arise from displacement of weather events due to the high dimensionality of the NWP model. If a small-scale weather event that causes a strong positive temperature anomaly at 850 hPa, for instance deep moist convection, occurs in one simulation, that event is unlikely to occur at the exactly same grid point at the same time in another simulation. Figure 6.11 demonstrates how intermittency and small scale of such events affect correlations. North of the ”A” label in Fig. 6.11a, a positive temperature anomaly within a convective updraft is found and enhanced by adding the PSP scheme (Figure 6.11b). This is a particular case in which the displacement does not happen, and with which we can directly understand the impact as an amplitude. However, by further adding MPP to that, the convection is displaced and shows paired positive and negative impacts due to displacement (Fig. 6.11c). The next example, convection to the south of the sign ”B” indicates that missing convection shows strong sign-reversed impact between

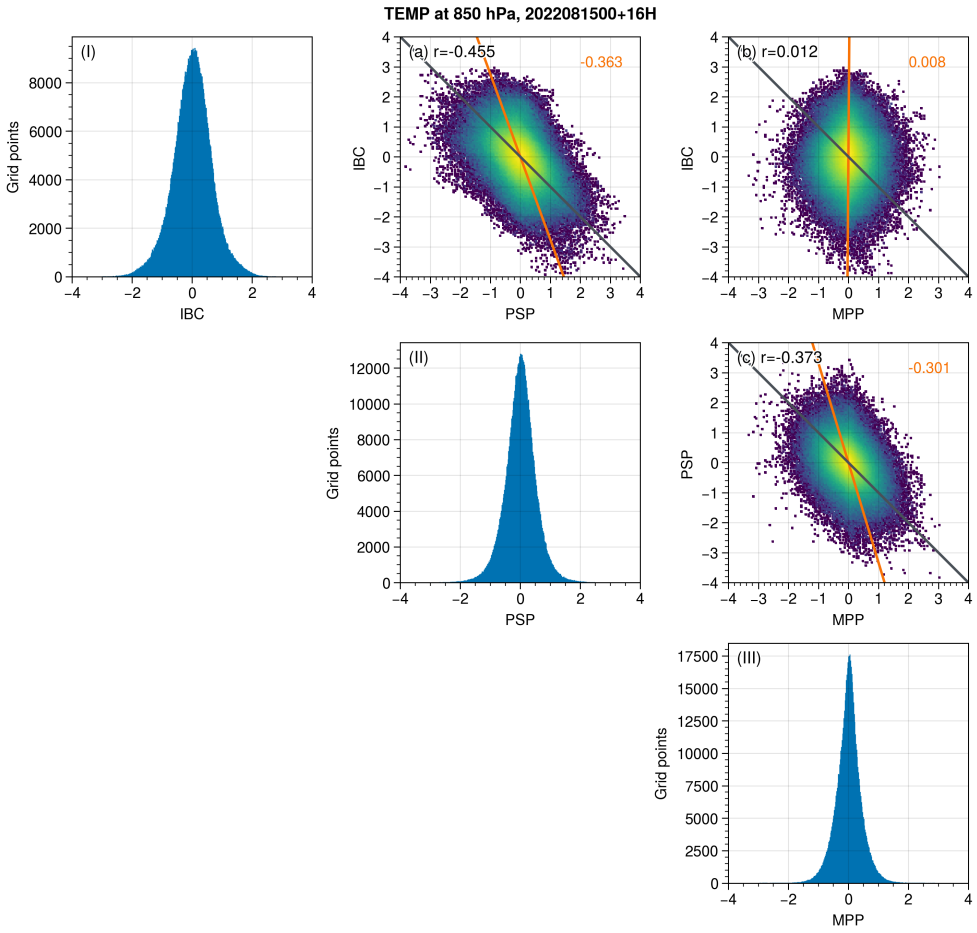

Figure 6.10: Cross correlation of uncertainty impacts for temperature at 850 hPa during peak convection for weak forcing case (1600 UTC of 15 August) aggregated across the Germany domain. Diagonal panels show impact sample distributions for (I) IBC, (II) PSP, and (III) MPP. Off-diagonal panels shows scatter densities of (a) IBC-PSP, (b) IBC-MPP, and (c) PSP-MPP impacts. Correlation coefficient values r are in the top left corner. Black lines indicate the perfect negative correlation line. Orange lines indicate linear regressions that estimate x-axis values from y-axis. The slope of the regression line is written in orange text. Grid points where either the three variances are less than 0.1 K^2 are excluded.

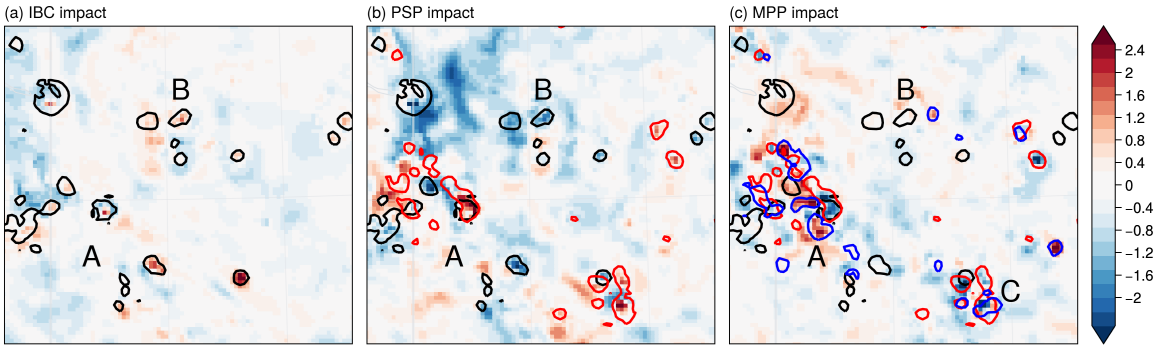

Figure 6.11: Example of temperature impact distributions for (a) IBC, (b) PSP, and (c) MPP at the 500 hPa pressure level at peak convection (1600 UTC on 15 August). An area of 180 x 180 km over northern Germany is shown. The colour unit is K. Black, red, and blue contours indicate location of convective updrafts (vertical wind > 1.0 m/s) of ensemble member 2 in the I, IP, and IPM ensembles, respectively. Refer to the text for letters A, B and C.

IBC and PSP (Fig. 6.11a,b) and neutral impact of MPP (Fig. 6.11c). This explains why only PSP responds to the IBC impact; the displacement is unlikely to survive two paths in the pathway. In a final example, PSP and MPP impact have an updraft at the same location (west of "C" in Fig. 6.11c) but still show the opposite sign of impact. This is a case that was caused by displacement in time. Despite the precise location of convection, the short duration of strong upward heat transport leads to the opposite sign of impact. Thus, such displacement has a strong effect as a factor pushing anomalous signals towards the ensemble mean, causing negative correlation between pre-existing and added impact. This "higher-order effect" is inherent in random disturbances of the atmospheric model state given by any kinds of perturbations. Although this method calculates grid-point correlations, the effect that the correlation shows is analogue to the spatial correlation concept of "correspondence in phase" in Murphy (1995). That is, negative correlation means opposite phase between impact and no correlation means that the phase is shifted by 1/4 wavelength, or completely randomly for continuous wavy fields like planetary-scale atmospheric motion. Since convective events distribute intermittently, phase shifts greater than 1/2 wavelength leads to perfect negative correlation in convection forecasts. However, if the higher-order effect and phase shifts were a particularly dominant factor, the correlation coefficient and regression slope would be close to -1. There are factors mitigating this effect, which will be discussed to answer the second question posed above.

Answering the latter question, what weakens the linear relationship given by the higher-order effect, is more challenging because it is a combination of several factors. Aligning with the idea of "correspondence in phase", we must ask what are factors leading to subtle phase shifts. One factor is that an anomalous impact typically combines contributions from multiple scales. Displacement only cancels out small-scale contributions, leaving the large-scale impact persists. Consequently, even if a strong localised anomaly is displaced,

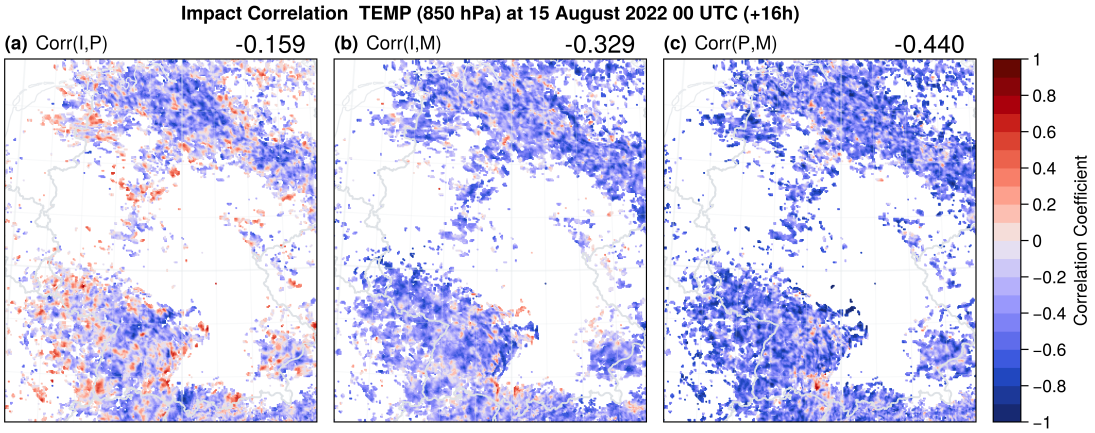

Figure 6.12: As for Fig. 6.8, but calculated within the M-path.

the neighbourhood may still be spatially or temporally correlated with the anomaly, mitigating the effect of displacement. Thus it appears as a form of spatial correlation within a neighbourhood, in line with the results in Minamide and Posselt (2022). Their work showed that convective moistening at the location of the convection peak is correlated with a neighbourhood with a radius of a few tens of kilometres. Another possible source is the nature of physical processes. In Fig. 6.10a, when the IBC impact produces a cool anomaly of around -2.0, the PSP impact is shown to be +2.0 or greater at many grid points. In contrast, when IBC has a strong warm anomaly of around +2.0, the PSP impact rarely shows values smaller than -2.0. This type of asymmetric response is more tied to specific physical processes, such as increased subsidence by the PSP scheme, which warms 850 hPa temperature in ensemble members with relatively less convection, but is unlikely to cool a warm ensemble member by suppressing convection. Last but not least, generation and prolongation of convection due to added model uncertainty schemes are new anomalous signals that cannot be explained as a phase shift of the pre-existing impact. The extent to which the negative correlation deviates from being perfect provides a measure of how much convection are modulated beyond simply being relocated from its original position. As this is the only controllable factor among those mentioned above, harnessing the model uncertainty scheme to generate and prolong convection, not merely displace it, is necessary to effectively increase ensemble variance.

6.4 Effect of pathway on correlation between impacts

Negative correlations between successively added impacts erode the effectiveness of adding uncertainty representations. This raises a question: does the order of adding representations alter the correlations? Figure 6.12 displays how changing the pathway modulates correlation fields. The IBC-PSP correlation shows a weaker average correlation (r = -0.159) and a noisy pattern in regions where convection is not very active at the time of

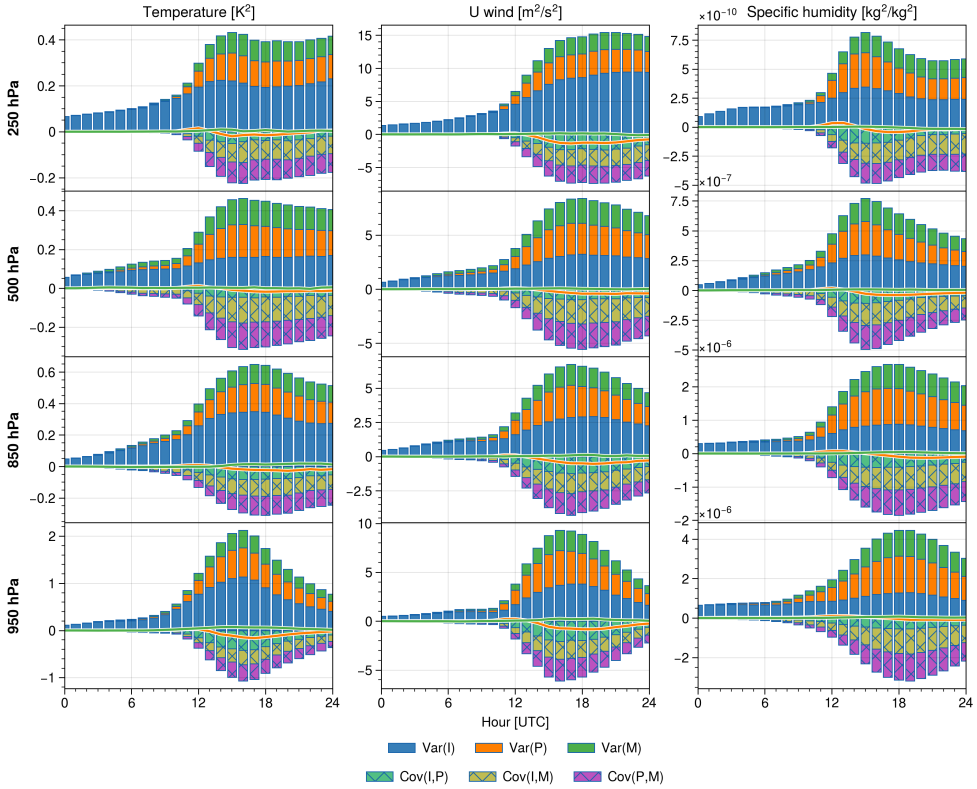

Figure 6.13: As for Fig. 6.4, but calculated within the M-path.

the display (Fig. 6.12a). However, systematic negative correlation is still present in the convectively active region (the north band in Fig. 5.8b). The correlation between IBC and MPP impacts becomes negative (r = -0.329) instead (Fig. 6.12a), while the MPP-PSP correlation remains almost unchanged (Fig. 6.12c). The implication is that the noisy part in Fig. 6.12a arises from the effect given by adding the impact of PSP on top of the pre-existing impact of MPP (the higher-order effect), while the systematic negative signal part shows inherent correlation between PSP and IBC. As expected from the previous discussions, the IBC-PSP correlation is also quite small during strong forcing due to the weak higher-order effect (not shown). This result aligns well with the intuition that the PSP scheme is typically not active in strong forcing conditions.

As the pathway changes the correlation structure significantly, does changing the pathway contribute to increasing the total variance? We know the answer is "no", from the fact that the IPM ensemble and its ensemble spread is unchanged for both pathways. However, we can understand why it happens. Figure 6.13 shows the evolution of variances and covariances within the M-path. The relative importance of respective covariance terms is very different to that from the P-path, but the individual variances remain unchanged. Again, adding both PSP and MPP does not produce a large increase in the total variance (orange and green lines). Therefore, changing the pathway only relocates the higher-order

Table 6.1: Average correlation coefficients for (rows) different pathways and (larger columns) forcing conditions. The bottom row shows absolute differences between the pathways. Values for averaged temperature correlation at 850 hPa between 1500 UTC and 2400 UTC are shown.

	Weak			Strong		
	IP	IM	PM	IP	IM	PM
P-path	-0.383	0.009	-0.361	-0.343	-0.016	-0.365
M-path	-0.146	-0.269	-0.389	-0.122	-0.267	-0.382
|P-path - M-path|	0.237	0.278	0.028	0.221	0.251	0.017

effect and barely alters individual variances and total variance. Assessing different pathways is like watching the same impact from different angles, meaning that a single pathway is sufficient to understand impact evolution and their interactions.

Despite the redundancy in examining different pathways for understanding the combined impact of uncertainty representations, comparing correlation coefficients between pathways allows the estimation of the saturation correlation due to higher-order effects. Table 6.1 shows the difference in correlation coefficients between pathways. The correlations between IBC and PSP, IBC and MPP change by approximately 0.25 for weak forcing, which represents the correlation coefficient given by the higher-order effect. In contrast, the correlation between PSP and MPP is barely affected by the pathway because PSP and MPP are added successively in both pathways. If other uncertainties are inserted in between, they will show weaker correlation. The differences between the pathways are slightly larger in weak forcing regimes, as the effect of displacement is smaller in strong forcing conditions.

Assuming this derived difference value is valid for any impact combinations, it allows for a rough estimate of the net gain in total variance only including the higher-order correlation. Consider a case of variance given by added uncertainty Var_{add} and ensemble variance without the added one Var_{pre}, with their correlation being $-\gamma$. Using Eq. 2.13, the criterion for the total variance Var_{total} to increase is:

$$\begin{aligned} Var_{total} - Var_{pre} = Var_{add} + 2\sqrt{Var_{pre}}\sqrt{Var_{add}}(-\gamma) > 0 \\ \sqrt{Var_{add}}/\sqrt{Var_{pre}} - 2\gamma > 0 \\ Var_{add}/Var_{pre} > 4\gamma^2. \end{aligned} \tag{6.5}$$

thus, to achieve a net gain in total variance after correlation saturation, the variance due to the added uncertainty needs to be larger than $4\gamma^2$. This corresponds to a fourfold of the pre-existing variance when $\gamma = 0.25$. If variance due to an added uncertainty representation is expected to be below this level, that representation will not contribute to increasing ensemble variability. However, caution is required as this threshold value varies for different variables and levels.

Although individual correlation structures largely depend on pathways, variance and net gain in total variance do not. The important message we can get from these results is

that once robust statistics for the higher-order effect are obtained, it is possible to roughly estimate whether an added uncertainty representation could increase total variance or not, without screening multiple pathways. Ideally, these statistics should be derived considering flow dependence to estimate added values for bespoke events.

6.5 Summary and Conclusions

This section studies the budget of variability, exploring **how forecast uncertainties evolve and interact** using the novel method of decomposing total ensemble variance into a summation of variances from individual uncertainty sources and their correlations. Their flow-dependent behaviour is examined by classifying two weakly forced and seven strongly forced days from a month-long ensemble dataset. The ensemble design follows a simplified Morris concept (described in Section 2.5) and includes a sufficient sub-ensemble size to sample the impact of model uncertainties, as discussed in Chapter. 5. Initially, an ensemble with perturbed IBCs is created. Then, model uncertainty representations, including PSP and MPP, are added sequentially. To investigate the effect of the order in which uncertainties are introduced, different sequences of adding these representations are tested, creating a cyclic graph of pathways.

The evolution of forecast uncertainties from various sources exhibits distinct behaviours across variables, heights, and convective regimes. Although individual uncertainty contributions can be substantial, their combined impact on the total ensemble variance is often limited due to negative correlations arising from the displacement of weather systems. This inevitable higher-order effect plays a crucial role in determining the net variance increase when incorporating new uncertainty representations. However, the strength of these negative correlations tends to saturate, suggesting a potential for increasing ensemble spread by introducing additional uncertainty sources once the correlation saturates, in a statistical sense.

The spatial patterns of uncertainty variances and correlations reveal intricate relationships with convective activity and model uncertainty behaviour. Both PSP and MPP actively insert variance within subregions of IBC activity where local convection is active. This strong link between local convection and model uncertainty is evident in weakly forced conditions. Conversely, in strongly forced conditions, the large-scale forcing dominates, leaving less room for model uncertainties to further modulate convection. Although this overlap of active regions shows some redundancy in model uncertainty representations for grid-point variance, they insert variability at finer scales than the IBC uncertainty, aligning with previous studies e.g. Hermoso et al. (2021); Chen et al. (2024).

While the pathway of adding uncertainty representations is crucial for determining individual correlations between impact, it does not change the variances introduced by the uncertainty sources. Performing ensemble simulations along a single pathway would suffice to verify if a new representation adds variability or not. The multi-pathway approach would be useful to disentangle the higher-order effect and inherent correlation from the combined correlation derived within a single pathway. Quantifying the correlations in-

troduced by the higher-order effect allows estimating the potential variance increase from adding new uncertainty representations. To achieve a net gain in total variance after correlation saturation, the variance due to the added uncertainty needs to be larger than a threshold determined by the higher-order correlation.

Although additional model uncertainty representations may have a marginal direct impact on increasing the total variance of these variables, it does not preclude beneficial impacts on other metrics. Stochastic drift (Romine et al., 2014; Berner et al., 2017) could alter the mean state of ensemble forecasts as a result of accounting for previously missed uncertain processes. This can lead to systematically reduce biases and eventually improve the overall reliability of the forecasts (e.g. McTaggart-Cowan et al., 2022; Puh et al., 2023).

This study gives some indication of which combination of model uncertainty representations are effective or redundant. However, it is important to acknowledge the limitations of this work. The current analysis focuses specifically on grid-point ensemble variance for free-atmospheric variables over a one-month period during summer. Extending the investigation to other regions and seasons will provide a more comprehensive view across diverse weather situations. Variables that increases in variance with accumulated values, such as precipitation, may exhibit different responses. Looking at more than a grid point is also important. Looking beyond a single grid is also important. Scale-dependent metrics (e.g. FSS (Roberts and Lean, 2008), Structure-Amplitude-Location scores (Wernli et al., 2008), and spectral analyses (Chen et al., 2024)) will help account for the effect of displacements and might reveal different budget evolution. Extending the method to these metrics is currently underway as a follow-up study to this thesis. Finally, we emphasise that it is essential to perform experiments using a full EPS in order to accurately account for the higher-order effect.

Chapter 7

Conclusions

Convective-scale EPSs typically underestimate forecast uncertainty because they do not adequately represent all sources of uncertainty in their model formulations. Accounting for these model uncertainties is crucial for achieving reliable ensemble forecasts. While a number of studies have investigated the individual impact of those uncertainties (e.g. Buizza et al., 1999; Leoncini et al., 2010; Berner et al., 2011b; Clark et al., 2016; Keil et al., 2019), the combined impact of multiple uncertainty sources has only recently attracted attention (e.g. Berner et al., 2015; Flack et al., 2021; Hermoso et al., 2021; Frogner et al., 2022). Even fewer studies have considered the influence of varying meteorological environmental conditions (Kühnlein et al., 2014; Flack et al., 2018; Puh et al., 2023; Chen et al., 2024). The purpose of this thesis is to add knowledge in this field by specifically investigating how the combined impact of model uncertainties evolves. The thesis is tracing the growth of variability, starting from daily and mesoscales and progressing to hourly and kilometre scales, under distinct atmospheric flow conditions. The thesis' overarching research question '**How different sources of uncertainty influence convective-scale forecast variability under varying convective forcing regimes?**' is explored from four different perspectives presented in separate chapters.

Two representations of model uncertainty were applied in a convection-permitting model. The first scheme, the microphysical parameter perturbations (MPP), perturbs CCN density and shape parameters of CDSD (Barthlott et al., 2022b). The second scheme, the physically-based stochastic parameterisation (PSP) scheme (Kober and Craig, 2016; Hirt et al., 2019), introduces feedback from subgrid-scale turbulence in the planetary boundary layer. Both schemes were introduced to the state-of-the-art convection-permitting model ICON-D2 (D. Reinert et al., 2021), in parallel with standard uncertainty representations of initial and lateral boundary conditions (IBC). The dependence of the impact on varying atmospheric situations was examined by classifying experiment days based on the convective adjustment time scale τ_c (Done et al., 2006). Convection is generally more unpredictable in weakly forced regimes, where poorly-represented local factors dominate the triggering of convection.

The first research question of **How does the relative impact of individual and combined uncertainties vary under different convective forcing regimes?** was

explored in Chapter 3. A combined ensemble of 20 IBC and 9 MPP members was extracted using a subsampling method to quantify the relative impacts of individual perturbations. The analysis showed that the ranges (90% confidence interval) of the relative MPP impact on 24-hour precipitation averaged over Germany were twice as sensitive in weak forcing compared to strong forcing. When MPP was combined with IBC perturbations, they elongated the tail of forecast variability without altering the distribution shape. This suggests a compensating effect that mitigates added variability, motivating further investigation of the combined variability when new uncertainty representations are added. Additionally, the sensitivity of the MPP impact to IBC highlights the necessity of using a full EPS for proper quantification of the impact of combined uncertainties.

The second question of **On what spatial scales does the uncertainties impact?** aimed to understand the scale-dependent impact. A three-month trial experiment including the PSP scheme (Puh et al., 2023) enabled systematic evaluation of the forecast error and spread in varying convective regimes. To quantify spatial variability of precipitation, the FSS technique (Roberts and Lean, 2008) and its variant spatial error and spread (Dey et al., 2014) were used. The study found that during weak forcing, the spatial error and spread of hourly precipitation strongly varied with the diurnal cycle. The displacement scale of error and spread during the convective period indicated spatial underdispersion of convective precipitation. Conversely, during strong forcing, the forecast error and spread increased steadily, exhibiting a better error-spread relationship. Activating the PSP scheme systematically increased spread on scales less than 50 km during weak forcing. The variability injected on the smallest scale during convection initiation evolved upscale until convection ceased. On scales larger than 50 km, the spread decreased as the PSP scheme facilitated convection initiation within a larger spatial window. Overall, adding the PSP scheme contributed to a better error-spread relationship across scales. This benefit was also found in strong forcing, but with a much smaller amplitude. Therefore, the PSP scheme selectively affected spatial variability during weak forcing, exhibiting the known upscale error growth behaviour (Zhang et al., 2007).

In a case study combining PSP and MPP, it was shown that their individual impact is additive. Although PSP and MPP perturbations showed similar up-scale growth in spatial variability, their impact acted on different spatial scales when combined with IBC uncertainty. The combined impact appeared to be the sum of the scale-dependent impact of the PSP scheme and the scale-independent impact of MPP. This suggests that PSP and MPP introduce spatial variability through different mechanisms. The PSP scheme primarily initiates and relocates convection, whereas MPP changes the amplitude of convective precipitation and subsequently alters location of strong precipitation signals. Although the impact of adding MPP was mostly compensated for with respect to the area-averaged precipitation amount (see Chapter 3), the impacts on spatial variability were additive. This finding led to the development of a framework to understand the budget of variability, exploring statistically how variances evolve and interact over space and time.

Motivated by these results, a novel framework named the "variability budget" has been developed to quantify the variability from various uncertainty sources and their interactions. This method decomposes total ensemble variance into the sum of individual

variances and their correlation terms, enabling the quantification of the effect of a certain model uncertainty representation on sampling of forecast variability as well as the diagnosis of the efficiency of a scheme to increase total forecast variability. The third research question of **Does model uncertainty need as many ensemble members as IBC to accurately estimate variability?** is addressed in Chapter 5 building on this new framework.

The convergence of variance and correlation with increasing ensemble size n of both IBCs and random seeds (RSD) in the PSP scheme is examined. Increasing nIBC continuously reduced the sampling uncertainty of the total variance and the individual variances attributed to IBC and PSP, following the $n^{-1/2}$ slope as shown in Craig et al. (2022); Tempest (2023). In contrast, increasing nRSD barely contributed to a reduction in sampling uncertainty. The variability stemming from the PSP scheme is partly covered by that from IBC, so smaller nRSD is needed to let the PSP impact converge when nIBC is sufficiently large ($n > 10$). Even a single random seed per IBC already seems suitable.

The impacts of IBC and PSP were found to be negatively correlated. The sampling error of correlation coefficient was computed using bootstrapping and compared to the theoretical convergence rate derived from the Fisher-z transform (Fisher, 1921). The sampling uncertainty of the correlation was reduced by increasing either nIBC or nRSD. However, the computed sampling uncertainty was systematically larger than the Fisher-z estimate. In addition, the sampling uncertainty decreased more slowly with increasing nRSD, indicating an overestimation of the effect of nRSD by the theory. This also suggests the superiority of nIBC to nRSD to achieve a better sampling of variability. Thus, the convergence of the impact correlation in the NWP application does not follow the Fisher-z theory.

In Chapter 6, the evolution of variances from different uncertainty sources and their interactions were traced from their origin with addressing the fourth research question of this thesis: **How does forecast uncertainty from different sources evolve and interact over time?** ICON-D2 ensemble simulations covering one entire summer month were performed to effectively sample the impact given by adding PSP and MPP to IBC perturbations, based on the findings in Chapter 5. The results demonstrated that the added uncertainty representations are actually acting although they seem to have almost no impact on the total (combined) variance. It is found that their impacts are negatively correlated with the pre-existing impact. This negative correlation is mainly caused by the higher-order effect arising from displacements of convective cells that foster the growth of variability and cancel out the ingested additional impact on the total variance. The model uncertainty variance grows similarly in both convective forcing regimes, but with lower relative importance of model uncertainty during strong forcing. Interestingly, the negative correlations saturate at some point suggesting that a quick saturation of correlation combined with continuing variance growth could lead to a net variance increase. At first sight this cancellation due to the correlation appears to contradict the additive impact found in Chapter 4. However, the scale-dependent metric takes into account the effect of spatial displacements and shows the impact from a different perspective of forecast variability.

The closer inspection of variance and correlation structures demonstrated that both

PSP and MPP insert variance in regions of large IBC spread. Displacements of weather phenomena in these regions, induced by the high dimensionality of the NWP model, were accounted for as a redundant effect, leading to the systematic negative correlation between the impacts. To increase total variance on a grid-point scale, a scheme needs to overcome the barrier posed by this effect of displacements. Comparing the correlation structures between different pathways of successively adding uncertainty representations disentangled the effect of displacements from the correlation found in a single pathway. Obtaining robust statistics on this effect can help to estimate the effectiveness of a new uncertainty scheme in increasing total variance.

Finally, the added knowledge for improving forecast reliability obtained by answering the overarching question of **How different sources of uncertainty influence convective-scale forecast variability under varying convective forcing regimes?** is summarised as follows: Different aspects of model uncertainty impact were precisely quantified by employing suitable metrics for each aspect. A new method designed to specifically assess the combined impact of uncertainties was proposed. The strong dependence of model uncertainty impact on IBC was objectively measured, with the impact assessed in presence of IBC uncertainty throughout the thesis. The significant sensitivity of moist convection to model uncertainty was visualised. Convective activity was a key factor in the growth of forecast uncertainties, regardless of the processes directly influenced by model uncertainty schemes. The importance of flow-dependent evaluation based on the different characteristics of convection is further emphasised, with the findings that the impact of model uncertainty was more pronounced in weak forcing regimes irrespective of the metrics used for measurement.

Although this thesis highlighted various aspects of the combined model uncertainty impact, there are some limitations. First, the research was performed only for summer convection over central Europe. The model uncertainty impact during winter, when large-scale baroclinic forcing is more dominant, and in monsoonal regions, where oceans modulate convection, could be different. Second, the analyses were based on aggregated statistics rather than process-level inspections due to the lack of computational capability to store high-resolution datasets required for such detailed analysis. Those analysis, like tracking individual convective cells, could provide additional insights into physical mechanisms driving the evolution of variability. Another caveat is the non-equivalent design of the PSP scheme and MPP. This made it challenging to isolate the contributions of homogeneous parameter structures (in MPP) and randomly generated spatial patterns (in PSP) to the differences found between their impacts. The use of a stochastic approach for MPP could significantly change the relative contributions of PSP and MPP impacts, as suggested by other studies Christensen et al. (2015); Stanford et al. (2019); McTaggart-Cowan et al. (2022).

The findings of this thesis open up many potential topics for future research. First, applying the methods used in this study to other uncertainty representations will facilitate a deeper understanding of variability interactions. Second, extending the variability budget method to handle either quantiles of a distribution or multi-modal distributions will significantly broaden its usability. This extension would enable quantifying the evolu-

tion of extreme values, which is particularly crucial for mitigating potential socioeconomic impacts caused by extreme weather events. Additionally, the variability budget method can be applied to spatial variability measures. Since many scale-dependent metrics employ squared differences between ensemble members in their formulations, the variability budget method can be readily extended to cover such metrics. A follow-up study is currently underway, applying the budget analysis to scale-dependent metrics for evaluating precipitation variability to understand the additive impact of model uncertainties on these metrics found in Chapter 4 of this thesis.

List of Abbreviations

Abbreviation	Description
CAPE	Convective Available Potential Energy
CCN	Cloud Condensation Nuclei
CDSD	Cloud Droplet Size Distribution
CI	Confidence Interval
CONUS	CONtinental United States
DA	Data Assimilation
DS	Displacement Scale
DWD	Deutscher WetterDienst
ECMWF	European Centre for Medium-range Weather Forecasts
EPS	Ensemble Prediction System
FSS	Fractions Skill Score
IBC	Initial and Boundary Condition
IC	Initial Condition
ICON	ICOsahedral Non-hydrostatic
KENDA	Kilometer-scale ENsemble Data Assimilation
LBC	Lateral Boundary Condition
LETKF	Local Ensemble Transform Kalman Filter
MPP	Microphysical Parameter Perturbations
NWP	Numerical Weather Prediction
PDF	Probability Density Function

PSP	Physically based Stochastic Perturbations
RGE	Regional Grand Ensemble
RSD	Random SeeD
SPP	Stochastic Parameter Perturbations
SPPT	Stochastically Perturbed Parametrisation Tendency
UTC	Coordinated Universal Time

List of Figures

List of Tables

Bibliography

Awad, M., Senga Kiesse, T., Assaghir, Z. and Ventura, A. (2019) Convergence of sensitivity analysis methods for evaluating combined influences of model inputs. *Reliability Engineering & System Safety*, **189**, 109–122.

Bachmann, K., Keil, C., Craig, G. C., Weissmann, M. and Welzbacher, C. A. (2020) Predictability of Deep Convection in Idealized and Operational Forecasts: Effects of Radar Data Assimilation, Orography, and Synoptic Weather Regime. *Mon. Wea. Rev.*, **148**, 63–81.

Barthlott, C. and Hoose, C. (2018) Aerosol effects on clouds and precipitation over central Europe in different weather regimes. *J. Atmos. Sci.*, **75**, 4247–4264.

Barthlott, C., Zarboo, A., Matsunobu, T. and Keil, C. (2022a) Impacts of combined microphysical and land-surface uncertainties on convective clouds and precipitation in different weather regimes. *Atmos. Chem. Phys.*, **22**, 10841–10860.

— (2022b) Importance of aerosols and shape of the cloud droplet size distribution for convective clouds and precipitation. *Atmos. Chem. Phys.*, **22**, 2153–2172.

Baur, F., Keil, C. and Barthlott, C. (2022) Combined effects of soil moisture and microphysical perturbations on convective clouds and precipitation for a locally forced case over Central Europe. *Q. J. R. Meteorol. Soc.*

Baur, F., Keil, C. and Craig, G. C. (2018) Soil moisture–precipitation coupling over Central Europe: Interactions between surface anomalies at different scales and the dynamical implication. *Q. J. R. Meteorol. Soc.*, **144**, 2863–2875.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Cámara, A., Christensen, H. M., Colangeli, M., Coleman, D. R., Crommelin, D., Dolaptchiev, S. I., Franzke, C. L., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., Palmer, T. N., Penland, C., Sakradzijaja, M., Von Storch, J. S., Weisheimer, A., Weniger, M., Williams, P. D. and Yano, J. I. (2017) Stochastic Parameterization: Toward a New View of Weather and Climate Models. *Bull. Am. Meteorol. Soc.*, **98**, 565–588.

Berner, J., Fossell, K. R., Ha, S.-Y., Hacker, J. P. and Snyder, C. (2015) Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations. *Monthly Weather Review*, **143**, 1295–1320.

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A. and Snyder, C. (2011a) Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. *Mon. Wea. Rev.*, **139**, 1972–1995.

— (2011b) Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. *Monthly Weather Review*, **139**, 1972–1995.

Bjerknes, V. (1904) Das problem der wettervorhersage, betrachtet vom standpunkte der mechanik und der physik. *Meteorol. Z.*, **21**, 1–7. (translated and edited by Volken E. and S. Brönnimann. – Meteorol. Z. 18 (2009), 663–667).

Blake, B. T., Carley, J. R., Alcott, T. I., Jankov, I., Pyle, M. E., Perfater, S. E. and Albright, B. (2018) An Adaptive Approach for the Calculation of Ensemble Gridpoint Probabilities. *Wea. Forecasting*, **33**, 1063–1080.

Bonavita, M., Hólm, E., Isaksen, L. and Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **142**, 287–303.

Bouttier, F., Vié, B., Nuissier, O. and Raynaud, L. (2012) Impact of Stochastic Physics in a Convection-Permitting Ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721.

Buizza, R. and Leutbecher, M. (2015) The forecast skill horizon. *Q. J. R. Meteorol. Soc.*, **141**, 3366–3382.

Buizza, R., Miller, M. and Palmer, T. N. (1999) Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2887–2908.

Campolongo, F., Cariboni, J. and Saltelli, A. (2007) An effective screening design for sensitivity analysis of large models. *Environ. Modell. Softw.*, **22**, 1509–1518.

Casati, B., Dorninger, M., Coelho, C. A., Ebert, E. E., Marsigli, C., Mittermaier, M. P. and Gilleland, E. (2022) The 2020 International Verification Methods Workshop Online: Major Outcomes and Way Forward. *Bull. Am. Meteorol. Soc.*, **103**, E899–E910.

Charney, J. G. (1951) Dynamic Forecasting by Numerical Process. In *Compendium of Meteorology: Prepared under the Direction of the Committee on the Compendium of Meteorology* (eds. H. R. Byers, H. E. Landsberg, H. Wexler, B. Haurwitz, A. F. Spilhaus, H. C. Willett, H. G. Houghton and T. F. Malone), 470–482. Boston, MA: American Meteorological Society.

Charney, J. G., FjÖrtoft, R. and Neumann, J. V. (1950) Numerical Integration of the Barotropic Vorticity Equation. *Tellus*, **2**, 237–254.

Chen, I.-H., Berner, J., Keil, C., Kuo, Y.-H. and Craig, G. (2023) Classification of Warm-Season Precipitation in High-Resolution Rapid Refresh (HRRR) Model Forecasts over the Contiguous United States. *Monthly Weather Review*, **152**, 187–201.

Chen, I.-H., Berner, J., Keil, C., Thompson, G., Kuo, Y.-H. and Craig, G. (2024) To which degree do the details of stochastic perturbation schemes matter for convective-scale and mesoscale error growth? **under review**.

Christensen, H. M., Moroz, I. M. and Palmer, T. N. (2015) Stochastic and Perturbed Parameter Representations of Model Uncertainty in Convection Parameterization*. *J. Atmos. Sci.*, **72**, 2525–2544.

Cintineo, R., Otkin, J. A., Xue, M. and Kong, F. (2014) Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic GOES-13 Satellite Observations. *Mon. Wea. Rev.*, **142**, 163–182.

Clark, P., Roberts, N., Lean, H., Ballard, S. P. and Charlton-Perez, C. (2016) Convection-permitting models: A step-change in rainfall forecasting. *Meteorol. Appl.*, **23**, 165–181.

Clark, P. A., Halliwell, C. E. and Flack, D. L. (2021) A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part I: Formulation and Evaluation in a Convection-Permitting Model. *J. Atmos. Sci.*, **78**, 727–746.

Covey, C., Lucas, D. D., Tannahill, J., Garaizar, X. and Klein, R. (2013) Efficient screening of climate model sensitivity to a large number of perturbed input parameters. *Journal of Advances in Modeling Earth Systems*, **5**, 598–610.

Craig, G., Fraedrich, K., Lin, J., Monkrieff, M., Plant, R., Thomson, D. and Tompkins, A. (2005) Working group 1 report: Issues in convection.

Craig, G. C., Keil, C. and Leuenberger, D. (2012) Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Q. J. R. Meteorol. Soc.*, **138**, 340–352.

Craig, G. C., Puh, M., Keil, C., Tempest, K., Necker, T., Ruiz, J., Weissmann, M. and Miyoshi, T. (2022) Distributions and convergence of forecast variables in a 1,000-member convection-permitting ensemble. *Q. J. R. Meteorol. Soc.*, **148**, 2325–2343.

D. Reinert, F. Prill, H. Frank, M. Denhard, M. Baldauf, C. Schraff, C. Gebhardt, C. Marsigli and G. Zängl (2021) DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System.

Dagan, G. and Stier, P. (2020) Ensemble daily simulations for elucidating cloud-aerosol interactions under a large spread of realistic environmental conditions. *Atmos. Chem. Phys.*, **20**, 6291–6303.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. and Meester, L. E. (2005) *A Modern Introduction to Probability and Statistics*. Springer Texts in Statistics. London: Springer.

Déqué, M. (2011) Deterministic Forecasts of Continuous Variables. In *Forecast Verification*, chap. 5, 77–94. John Wiley & Sons, Ltd.

Deutscher Wetterdienst (DWD) (2022) Wetter und Klima - Deutscher Wetterdienst - Leistungen - Analysen radarbasierter stündlicher (RW) und täglicher (SF) Niederschlagshöhen. https://www.dwd.de/DE/leistungen/radolan/radolan.html.

Dey, S. R., Leoncini, G., Roberts, N. M., Plant, R. S. and Migliorini, S. (2014) A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107.

Done, J. M., Craig, G. C., Gray, S. L. and Clark, P. A. (2012) Case-to-case variability of predictability of deep convection in a mesoscale model. *Q. J. R. Meteorol. Soc.*, **138**, 638–648.

Done, J. M., Craig, G. C., Gray, S. L., Clark, P. A. and Gray, M. E. B. (2006) Mesoscale simulations of organized convection: Importance of convective equilibrium. *Q. J. R. Meteorol. Soc.*, **132**, 737–756.

Ebert, E. E. (2009) Neighborhood Verification: A Strategy for Rewarding Close Forecasts. *Wea. Forecasting*, **24**, 1498–1510.

Epstein, E. S. (1969) Stochastic dynamic prediction. *Tellus*, **21**, 739–759.

Fédou, J.-M. and Rendas, M.-J. (2015) Extending Morris method: Identification of the interaction graph using cycle-equitable designs. *Journal of Statistical Computation and Simulation*, **85**, 1398–1419.

Fisher, R. A. (1921) On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, **1**, 3–32.

Flack, D. L., Gray, S. L. and Plant, R. S. (2019) A simple ensemble approach for more robust process-based sensitivity analysis of case studies in convection-permitting models. *Q. J. R. Meteorol. Soc.*, **145**, 3089–3101.

Flack, D. L., Gray, S. L., Plant, R. S., Lean, H. W. and Craig, G. C. (2018) Convective-scale perturbation growth across the spectrum of convective regimes. *Mon. Wea. Rev.*, **146**, 387–405.

Flack, D. L. A., Clark, P. A., Halliwell, C. E., Roberts, N. M., Gray, S. L., Plant, R. S. and Lean, H. W. (2021) A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part II: Perturbation Growth within a Superensemble Framework. *J. Atmos. Sci.*, **78**, 747–761.

Fleury, A., Bouttier, F. and Couvreux, F. (2022) Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction. *Q. J. R. Meteorol. Soc.*, **148**, 981–1000.

Fouladi, R. T. and Steiger, J. H. (2008) The Fisher Transform of the Pearson Product Moment Correlation Coefficient and Its Square: Cumulants, Moments, and Applications. *Communications in Statistics - Simulation and Computation*, **37**, 928–944.

Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Hämäläinen, K., Kauhanen, J., Ivarsson, K.-I. and Yazgi, D. (2022) Model Uncertainty Representation in a Convection-Permitting Ensemble—SPP and SPPT in HarmonEPS. *Monthly Weather Review*, **150**, 775–795.

Frogner, I.-L., Singleton, A. T., Køltzow, M. Ø. and Andrae, U. (2019) Convection-permitting ensembles: Challenges related to their design and use. *Q. J. R. Meteorol. Soc.*, **145**, 90–106.

Glassmeier, F. and Lohmann, U. (2018) Precipitation Susceptibility and Aerosol Buffering of Warm- and Mixed-Phase Orographic Clouds in Idealized Simulations. *Journal of the Atmospheric Sciences*, **75**, 1173–1194.

Hande, L. B., Engler, C., Hoose, C. and Tegen, I. (2016) Parameterizing cloud condensation nuclei concentrations during HOPE. *Atmos. Chem. Phys.*, **16**, 12059–12079.

Hermoso, A., Homar, V. and Plant, R. S. (2021) Potential of stochastic methods for improving convection-permitting ensemble forecasts of extreme events over the Western Mediterranean. *Atmos. Res.*, **257**, 105571.

Hirt, M., Rasp, S., Blahak, U. and Craig, G. C. (2019) Stochastic Parameterization of Processes Leading to Convective Initiation in Kilometer-Scale Models. *Mon. Wea. Rev.*, **147**, 3917–3934.

Hohenegger, C. and Schär, C. (2007) Predictability and Error Growth Dynamics in Cloud-Resolving Models. *J. Atmos. Sci.*, **64**, 4467–4478.

Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, **52**, 1–17.

Igel, A. L. and van den Heever, S. C. (2017a) The importance of the shape of cloud droplet size distributions in shallow cumulus clouds. Part I: Bin microphysics simulations. *J. Atmos. Sci.*, **74**, 249–258.

— (2017b) The importance of the shape of cloud droplet size distributions in shallow cumulus clouds. Part II: Bulk microphysics simulations. *J. Atmos. Sci.*, **74**, 259–273.

Iooss, B. and Lemaître, P. (2015) A review on global sensitivity analysis methods. In *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications* (eds. C. Meloni and G. Dellino). Springer.

Jankov, I., Beck, J., Wolff, J., Harrold, M., Olson, J. B., Smirnova, T., Alexander, C. and Berner, J. (2019) Stochastically Perturbed Parameterizations in an HRRR-Based Ensemble. *Monthly Weather Review*, **147**, 153–173.

Ji, D., Dong, W., Hong, T., Dai, T., Zheng, Z., Yang, S. and Zhu, X. (2018) Assessing Parameter Importance of the Weather Research and Forecasting Model Based On Global Sensitivity Analysis Methods. *Journal of Geophysical Research: Atmospheres*, **123**, 4443–4460.

Johnson, A. and Wang, X. (2020) Interactions between physics diversity and multiscale initial condition perturbations for storm-scale ensemble forecasting. *Mon. Wea. Rev.*, **148**, 3549–3565.

Kalnay, E. (2003) *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge: Cambridge University Press.

Keil, C., Baur, F., Bachmann, K., Rasp, S., Schneider, L. and Barthlott, C. (2019) Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Q. J. R. Meteorol. Soc.*, **145**, 3102–3115.

Keil, C., Chabert, L., Nuissier, O. and Raynaud, L. (2020) Dependence of predictability of precipitation in the northwestern Mediterranean coastal region on the strength of synoptic control. *Atmos. Chem. Phys.*, **20**, 15851–15865.

Keil, C. and Craig, G. C. (2011) Regime-dependent forecast uncertainty of convective precipitation. *Meteorol. Z.*, **20**, 145–151.

Keil, C., Heinlein, F. and Craig, G. C. (2014) The convective adjustment time-scale as indicator of predictability of convective precipitation. *Q. J. R. Meteorol. Soc.*, **140**, 480–490.

Kober, K. and Craig, G. C. (2016) Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.*, **73**, 2893–2911.

Kühnlein, C., Keil, C., Craig, G. C. and Gebhardt, C. (2014) The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Q. J. R. Meteorol. Soc.*, **140**, 1552–1562.

Kunz, M., Abbas, S. S., Bauckholt, M., Böhmländer, A., Feuerle, T., Gasch, P., Glaser, C., Groß, J., Hajnsek, I., Handwerker, J., Hase, F., Khordakova, D., Knippertz, P., Kohler, M., Lange, D., Latt, M., Laube, J., Martin, L., Mauder, M., Möhler, O., Mohr, S., Reitter, R. W., Rettenmeier, A., Rolf, C., Saathoff, H., Schrön, M., Schütze, C., Spahr, S., Späth, F., Vogel, F., Völksch, I., Weber, U., Wieser, A., Wilhelm, J., Zhang, H. and Dietrich, P. (2022) Swabian MOSES 2021: An interdisciplinary field campaign for investigating convective storms and their event chains. *Frontiers in Earth Science*, **10**.

Lang, S. T. K., Lock, S.-J., Leutbecher, M., Bechtold, P. and Forbes, R. M. (2021) Revision of the Stochastically Perturbed Parametrisations model uncertainty scheme in the Integrated Forecasting System. *Quarterly Journal of the Royal Meteorological Society*, **147**, 1364–1381.

Laprise, R. (2003) Resolved Scales and Nonlinear Interactions in Limited-Area Models. *Journal of the Atmospheric Sciences*, **60**, 768–779.

Leith, C. E. (1974) Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Leoncini, G., Plant, R. S., Gray, S. L. and Clark, P. A. (2010) Perturbation growth at the convective scale for CSIP IOP18. *Q. J. R. Meteorol. Soc.*, **136**, 653–670.

Lorenz, E. N. (1963) Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, **20**, 130–141.

— (1969a) The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

— (1969b) Three Approaches to Atmospheric Predictability. *Bull. Amer. Meteor. Soc.*, **50**, 345–349.

Lynch, P. (2008) The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227**, 3431–3444.

Marinescu, P. J., Van Den Heever, S. C., Heikenfeld, M., Barrett, A. I., Barthlott, C., Hoose, C., Fan, J., Fridlind, A. M., Matsui, T., Miltenberger, A. K., Stier, P., Vie, B., White, B. A. and Zhang, Y. (2021) Impacts of Varying Concentrations of Cloud Condensation Nuclei on Deep Convective Cloud Updrafts—A Multimodel Assessment. *J. Atmos. Sci.*, **78**, 1147–1172.

Matsunobu, T., Keil, C. and Barthlott, C. (2022) The impact of microphysical uncertainty conditional on initial and boundary condition uncertainty under varying synoptic control. *Weather Clim. Dyn.*, **3**, 1273–1289.

Matsunobu, T., Puh, M. and Keil, C. (2024) Flow- and scale-dependent spatial predictability of convective precipitation combining different model uncertainty representations. *Quarterly Journal of the Royal Meteorological Society*, **150**, 2364–2381.

McTaggart-Cowan, R., Separovic, L., Charron, M., Deng, X., Gagnon, N., Houtekamer, P. L. and Patoine, A. (2022) Using Stochastically Perturbed Parameterizations to Represent Model Uncertainty. Part II: Comparison with Existing Techniques in an Operational Ensemble. *Mon. Wea. Rev.*, **150**, 2859–2882.

Minamide, M. and Posselt, D. J. (2022) Using Ensemble Data Assimilation to Explore the Environmental Controls on the Initiation and Predictability of Moist Convection. *Journal of the Atmospheric Sciences*, **79**, 1151–1169.

Mittermaier, M. and Roberts, N. (2010) Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score. *Wea. Forecasting*, **25**, 343–354.

Mittermaier, M., Roberts, N. and Thompson, S. A. (2013) A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol. Appl.*, **20**, 176–186.

Morales, A., Posselt, D. J., Morrison, H. and He, F. (2019) Assessing the Influence of Microphysical and Environmental Parameter Perturbations on Orographic Precipitation. *Journal of the Atmospheric Sciences*, **76**, 1373–1395.

Morris, M. D. (1991) Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, **33**, 161–174.

Murphy, A. H. (1995) The Coefficients of Correlation and Determination as Measures of performance in Forecast Verification. *Wea. Forecasting*, **10**, 681–688.

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M. and Serafin, S. (2024) The fractions skill score for ensemble forecast verification. *Q. J. R. Meteorol. Soc.*, **under review**.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J. and Sandu, I. (2017) Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, **143**, 408–422.

Palmer, T. and Hagedorn, R. (eds.) (2006) *Predictability of Weather and Climate*. Cambridge: Cambridge University Press.

Palmer, T., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J. (1992) Ensemble prediction. *ECMWF Technical memorandum*, 43.

Plant, B., Hally, A., Lock, S.-J., Ahlgrimm, M., Arpargus, M., Bauer, W., Cafaro, C., Chen, I., Fannon, J., Feddersen, H., Fleury, A., Frogner, I.-L., Gebhardt, C., Hamalainen, K., Hieronymus, M., Kauhanen, J., Kouhen, S., Kuntze, P., Lean, H., Mamgain, A., McCabe, A., Milan, M., Petch, J., Puh, M., Schraff, C., Stirling, A., Tempest,

K., Tsiringakis, A., Weidle, F., Wimmer, M. and Xu, Z. (2022) Summary and Recommendations from Working Group 1: Model uncertainty representations in convection-permitting / shorter lead-time / limited-area ensembles. In *ECMWF Workshop: Model Uncertainty*.

Plant, R. S. and Craig, G. C. (2008) A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics. *J. Atmos. Sci.*, **65**, 87–105.

Posselt, D. J., He, F., Bukowski, J. and Reid, J. S. (2019) On the Relative Sensitivity of a Tropical Deep Convective Storm to Changes in Environment and Cloud Microphysical Parameters. *Journal of the Atmospheric Sciences*, **76**, 1163–1185.

Puh, M., Keil, C., Gebhardt, C., Marsigli, C., Hirt, M., Jakub, F. and Craig, G. C. (2023) Physically based stochastic perturbations improve a high-resolution forecast of convection. *Q. J. R. Meteorol. Soc.*

Rasp, S., Selz, T. and Craig, G. C. (2018) Variability and Clustering of Midlatitude Summertime Convection: Testing the Craig and Cohen Theory in a Convection-Permitting Ensemble with Stochastic Boundary Layer Perturbations. *J. Atmos. Sci.*, **75**, 691–706.

Raynaud, L. and Bouttier, F. (2017) The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. *Q. J. R. Meteorol. Soc.*, **143**, 3037–3047.

Rezacova, D., Zacharov, P. and Sokol, Z. (2009) Uncertainty in the area-related QPF for heavy convective precipitation. *Atmos. Res.*, **93**, 238–246.

Richardson, L. F. (1922) *Weather Prediction by Numerical Process*. Cambridge Mathematical Library. Cambridge: Cambridge University Press, 2 edn.

Roberts, B., Clark, A. J., Jirak, I. L., Gallo, B. T., Bain, C., Flack, D. L. A., Warner, J., Schwartz, C. S. and Reames, L. J. (2023) Model Configuration versus Driving Model: Influences on Next-Day Regional Convection-Allowing Model Forecasts during a Real-Time Experiment. *Wea. Forecasting*, **38**, 99–123.

Roberts, N. M. and Lean, H. W. (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, **136**, 78–97.

Rodwell, M. J., Lang, S. T., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., Richardson, D. S. and Yamaguchi, M. (2016) Reliability in ensemble data assimilation. *Q. J. R. Meteorol. Soc.*, **142**, 443–454.

Romine, G. S., Schwartz, C. S., Berner, J., Fossell, K. R., Snyder, C., Anderson, J. L. and Weisman, M. L. (2014) Representing Forecast Error in a Convection-Permitting Ensemble System. *Monthly Weather Review*, **142**, 4519–4541.

Ryan, E., Wild, O., Voulgarakis, A. and Lee, L. (2018) Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. *Geosci. Model Dev.*, **11**, 3131–3146.

Scheck, L., Weissmann, M. and Bach, L. (2020) Assimilating visible satellite images for convective-scale numerical weather prediction: A case-study. *Q. J. R. Meteorol. Soc.*, **146**, 3165–3186.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Q. J. R. Meteorol. Soc.*, **142**, 1453–1472.

Schulz, J.-P., Vogel, G., Becker, C., Kothe, S., Rummel, U. and Ahrens, B. (2016) Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil. *Meteorologische Zeitschrift*, 607–620.

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J. and Coniglio, M. C. (2009) Next-Day Convection-Allowing WRF Model Guidance: A Second Look at 2-km versus 4-km Grid Spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.

Seifert, A. and Beheng, K. D. (2006) A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, **92**, 45–66.

Seifert, A., Köhler, C. and Beheng, K. D. (2012) Aerosol-cloud-precipitation effects over Germany as simulated by a convective-scale numerical weather prediction model. *Atmos. Chem. Phys.*, **12**, 709–725.

Selz, T. and Craig, G. C. (2015) Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe. *Mon. Wea. Rev.*, **143**, 813–827.

Shukla, J. (1998) Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science*, **282**, 728–731.

Shutts, G. and Pallarès, A. C. (2014) Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **372**, 20130284.

Skok, G. and Roberts, N. (2016) Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts. *Q. J. R. Meteorol. Soc.*, **142**, 2599–2610.

— (2018) Estimating the displacement in precipitation forecasts using the Fractions Skill Score. *Q. J. R. Meteorol. Soc.*, **144**, 414–425.

Slingo, J. and Palmer, T. (2011) Uncertainty in weather and climate prediction. *Philos. trans., Math. phys. eng. sci.*, **369**, 4751–67.

Sobol', I. M. (1990) On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe modelirovanie*, **2**, 112–118.

Stanford, M. W., Morrison, H., Varble, A., Berner, J., Wu, W., McFarquhar, G. and Milbrandt, J. (2019) Sensitivity of Simulated Deep Convection to a Stochastic Ice Microphysics Framework. *Journal of Advances in Modeling Earth Systems*, **11**, 3362–3389.

Tang, Y., Lean, H. W. and Bornemann, J. (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, **20**, 417–426.

Tempest, K. (2023) *Sampling uncertainty in ensemble forecasting: when do we have enough ensemble members?* Text.PhDThesis, Ludwig-Maximilians-Universität München.

Tempest, K. I., Craig, G. C. and Brehmer, J. R. (2023) Convergence of forecast distributions in a 100,000-member idealised convective-scale ensemble. *Q. J. R. Meteorol. Soc.*, **149**, 677–702.

Thompson, G., Berner, J., Frediani, M., Otkin, J. A. and Griffin, S. M. (2021) A Stochastic Parameter Perturbation Method to Represent Uncertainty in a Microphysics Scheme. *Mon. Wea. Rev.*, **149**, 1481–1497.

Vié, B., Nuissier, O. and Ducrocq, V. (2011) Cloud-Resolving Ensemble Simulations of Mediterranean Heavy Precipitating Events: Uncertainty on Initial Conditions and Lateral Boundary Conditions. *Mon. Wea. Rev.*, **139**, 403–423.

Vukicevic, T. and Errico, R. M. (1990) The Influence of Artificial and Physical Factors upon Predictability Estimates Using a Complex Limited-Area Model. *Monthly Weather Review*, **118**, 1460–1482.

Wellmann, C., I Barrett, A., S Johnson, J., Kunz, M., Vogel, B., S Carslaw, K. and Hoose, C. (2020) Comparing the impact of environmental conditions and microphysics on the forecast uncertainty of deep convective clouds and hail. *Atmos. Chem. Phys.*, **20**, 2201–2219.

Wernli, H., Paulat, M., Hagen, M., Frei, C., Wernli, H., Paulat, M., Hagen, M. and Frei, C. (2008) SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, **136**, 4470–4487.

White, B., Gryspeerdt, E., Stier, P., Morrison, H., Thompson, G. and Kipling, Z. (2017) Uncertainty from the choice of microphysics scheme in convection-permitting models significantly exceeds aerosol effects. *Atmos. Chem. Phys.*, **17**.

Wimmer, M., Raynaud, L., Descamps, L., Berre, L. and Seity, Y. (2022) Sensitivity analysis of the convective-scale AROME model to physical and dynamical parameters. *Q. J. R. Meteorol. Soc.*, **148**, 920–942.

Ying, Y. and Zhang, F. (2017) Practical and Intrinsic Predictability of Multiscale Weather and Convectively Coupled Equatorial Waves during the Active Phase of an MJO. *J. Atmos. Sci.*, **74**, 3771–3785.

Zacharov, P. and Rezacova, D. (2009) Using the fractions skill score to assess the relationship between an ensemble QPF spread and skill. *Atmos. Res.*, **94**, 684–693.

Zängl, G., Reinert, D., Rípodas, P. and Baldauf, M. (2015) The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.*, **141**, 563–579.

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C. C. (2007) Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**.

Zhang, Y., Fan, J., Li, Z. and Rosenfeld, D. (2021) Impacts of cloud microphysics parameterizations on simulated aerosol–cloud interactions for deep convective clouds over Houston. *Atmos. Chem. Phys.*, **21**, 2363–2381.

Zimmer, M., Craig, G. C., Keil, C. and Wernli, H. (2011) Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophys. Res. Lett.*, **38**, n/a–n/a.

Acknowledgements

First and foremost, I would like to express my gratitude to my Doktorvater Prof. Dr. George Craig and my project supervisor Dr. Christian Keil for their tremendous support and guidance throughout my PhD expedition. Their expertise, tenderness, patience, unlimited ideas, and constructive feedback have been absolutely invaluable in shaping my research and improving my skills required for growing up as a scientist. Furthermore, I would like to thank Prof. Dr. Markus Rapp for co-examining this thesis. I had the good fortune to have come across such great mentors in the early stage of my career.

I am grateful to my officemates and colleagues at MIM LMU who always gave me inspiration of research as well as feeling like home. Coffee break discussions and lunch meetings with the early career scientists enriched my research experience and helped me better organise myself. In addition, my research could not come to realisation without administrative and technical support from LMU. The infrastructure they have provided was supporting me to process everything necessary for completing PhD smoothly.

I would like to extend my thanks to the research programme "Waves to Weather" and my colleagues therein. The collaborations with them has been indispensable to make my project both enriching and intriguing. I am particularly thankful to Dr. Christian Barthlott for sharing knowledge in our joint project, Dr. Audine Laurian for her unwavering support in management and Dr. Christian Grams for having encouraged me to participate in the W2W. Every W2W meeting was motivating and helping build up my communication skill as well as sense of internationality. The supportive W2W community was a lifeline during the Covid-19 pandemic, especially while being in a country geographically and culturally distant from where I grew up. It is my honour to have been involved with such a world-leading research programme.

I also thank to nice collaborators in the Deutscher Wetterdienst. The data, model and more than all their productive feedback were essential for performing the research in the best environment for my research topic.

この博士論文はミュンヘンで過ごした 4 年間だけではなく、今まで得られたすべての学びの上に成り立っています。辛抱強く私に機会をくださった両親に心からの感謝を贈ります。思えば遠くまで来たものですが、これもひとえに両親が私の自由奔放な生き方を許容してくれたお陰です。さもなければ今の私は在りませんでした。妻の桃子はこの謝辞のために最適なフォントを見つけ出すという重要にして最後の仕上げをしてくれました。筑波大学で人生の転機を与えてくださった田中博先生と松枝未遠先生への感謝は言葉で表せません。学問は勿論のこと、教育者としての立ち振る舞い

や心構えを学びました。両先生のように大きな器を持つことが、研究者として、教育者として、そして人間としての目標です。また、未熟な私を研究アシスタントとして教育してくださった気象庁気象研究所の斉藤和雄さま、川畑拓矢さま、農研機構農業環境変動研究センターの飯泉仁之直さまをはじめ、さまざまにご協力くださった研究所のみなさまに感謝を申し上げます。これらの援助のうちどれが欠けてもこの博士論文は完成しませんでした。

Of course, thank many people who were not mentioned above but helped me in school, university, and somewhere else.

Zuletzt möchte ich mich bei der großartigen Stadt München bedanken. Ohne die wunderbare Atmosphäre, die Kultur, die Natur und das Bier in Bayern hätte ich diese Doktorarbeit nicht fertigstellen können. Ich bin sehr glücklich, meine Universitätsjahre hier verbracht zu haben.