Observation Influence and Imbalance in Convective-Scale Data Assimilation

Theresa Alina Diefenbach

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM

München 2024

Observation Influence and Imbalance in Convective-Scale Data Assimilation

Theresa Alina Diefenbach

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

vorgelegt von
Theresa Alina Diefenbach

München, den 23. Mai 2024

Erstgutachter: Prof. Dr. George Craig

Zweitgutachter: Prof. Dr. Martin Weissmann

Tag der mündlichen Prüfung: 22. Juli 2024

Zusammenfassung

Die numerische Wettervorhersage (NWV) ist ein Anfangswertproblem, bei dem die Genauigkeit der Anfangsbedingungen entscheidend für den Erfolg der Vorhersage ist. Zwar stehen Beobachtungsdaten zur Definition des Anfangszustands zur Verfügung, doch sind sie im Vergleich zu den Anforderungen des Wettervorhersagemodells unzureichend. Um diese Lücke zu schließen, integriert die Datenassimilation (DA) Beobachtungsdaten mit einer früheren Modellvorhersage, um Anfangsbedingungen zu erhalten, die die beste Schätzung des atmosphärischen Zustands unter Berücksichtigung aller verfügbaren Informationen darstellen. Diese beste Schätzung wird als Analyse bezeichnet. Die DA ist für die NWV von grundlegender Bedeutung, doch ihre Komplexität in Verbindung mit Annahmen, die in der Realität nicht immer erfüllt sind, kann zu Störungen der physikalischen Gleichgewichte in der Analyse führen, was die Wirksamkeit der DA einschränken kann. Daher ist die Überwachung des DA-Prozesses wichtig, insbesondere mit dem Aufkommen neuer und komplexer Beobachtungsdatenquellen, wie Satelliten oder bodengestützter Fernerkundung. Die Überwachung selbst stellt jedoch auch eine Herausforderung dar. In dieser Arbeit werden zwei neuartige Methoden zur Untersuchung der Effektivität eines konvektiv-skaligen DA-Systems vorgestellt. Dabei konzentrieren sich die Methoden vor allem auf den Einfluss der assimilierten Beobachtungen auf die Analyse und das mögliche Auftreten von Ungleichgewichten.

Im ersten Teil dieser Arbeit wird ein rechnerisch effizienter Ansatz zur Quantifizierung des Analyseeinflusses vorgestellt. Wir nutzen grundlegende Eigenschaften des Local Ensemble Transform Kalman Filters (LETKF), einer häufig verwendeten DA-Methode auf der konvektiven Skala, um die Analyseaktualisierung in Beiträge einzelner Beobachtungen zu zerlegen, die dann als partielle Analyse-Inkremente bezeichnet werden. Während diese Methode Elemente mit etablierten Analysesensitivitätsmaßen gemeinsam hat, bietet der vorgestellte Ansatz eine explizite und rechnerisch effizientere Möglichkeit, den Einfluss von Beobachtungen auf verschiedene Modellvariablen zu untersuchen. Darüber hinaus ermöglicht die neu entwickelte Methode die Identifizierung potenziell nachteiliger Beobachtungseinflüsse und erleichtert die Optimierung der DA-Einstellungen zur Verbesserung der Analysegenauigkeit.

Der zweite Teil dieser Dissertation befasst sich mit Diagnosetechniken zur Messung von ungewollten Ungleichgewichten in den durch Datenassimilation erhaltenen Analysezuständen. Diese ungewollten Ungleichgewichte beziehen sich häufig auf Störungen von

Kräftegleichgewichten, die normalerweise in der Atmosphäre vorherrschend sind. Die spezifische Form dieser Ungleichgewichte in Konvektions-auflösenden NWV-Modellen ist jedoch bisher nicht bekannt, und es wird erwartet, dass es sich von den Gleichgewichtsbedingungen in großskaligeren Modellen unterscheidet. Um dies zu untersuchen, werden drei verschiedene Metriken, die jeweils auf unterschiedlichen physikalischen Begründungen beruhen, implementiert und auf ein Konvektions-auflösendes NWV-System angewendet. Dazu gehört eine neu entwickelte Methode, die auf einem physikalischen Gleichgewichtsprinzip basiert, das auf der konvektiven Skala annähernd gültig ist, die sogenannte “Weak Temperature Gradient (WTG) Balance”, die im Rahmen dieser Arbeit zum ersten Mal implementiert und getestet wurde. Die Ergebnisse zeigen, dass die verschiedenen Metriken unterschiedliche Facetten des Ungleichgewichts zu erfassen scheinen, wobei sich die WTG-Metrik als besonders geeigneter Kandidat erweist, der komplementäre Ergebnisse zu bestehenden Methoden liefert.

Wir hoffen, dass die beiden vorgestellten Diagnoseinstrumente in Zukunft zu weiteren Fortschritten auf dem Gebiet der Datenassimilation beitragen und damit eine Verbesserung der Wettervorhersagegenauigkeit ermöglichen können.

Abstract

Numerical weather prediction (NWP) is an initial value problem where the accuracy of the initial conditions is critical to the success of the forecast. While observational data are available to define the initial state, they are insufficient compared to the requirements of the weather prediction model. To bridge this gap, data assimilation (DA) integrates observational data with a first-guess weather forecast to obtain initial conditions that represent the best estimate of the atmospheric state given all available information. This best estimate is commonly referred to as the analysis. DA is essential for NWP, however its complexity, coupled with assumptions that do not always reflect reality, may introduce noise in the form of imbalance, which has the potential to limit the effectiveness of DA. Therefore, monitoring the DA process is crucial, particularly with the advent of new and intricate observational data sources, such as satellites or ground-based remote sensing. Yet, monitoring itself presents challenges. This thesis presents two novel methods for investigating the effectiveness of a convective-scale DA system with respect to the incorporation of observational information into the analysis and the occurrence of possible imbalances due to DA.

In the first part of this thesis a computationally efficient approach to quantify the analysis influence of observations is presented. We use fundamental properties of the Local Ensemble Transform Kalman Filter (LETKF), a frequently used DA method in convective-scale DA, to disentangle the analysis update into the contributions of individual observations, which are called partial analysis increments. While this method shares elements with established analysis sensitivity measures, the presented approach offers a more explicit and computationally efficient way to explore the influence of observations across various model variables. Moreover, it enables the identification of potentially detrimental observation influence and facilitates the optimisation of DA settings for enhanced analysis accuracy.

The second part of this thesis deals with diagnostic techniques to assess spurious imbalances in the initial states obtained by data assimilation. This spurious imbalance often refers to a disturbance of the balance of forces that normally prevails in the atmosphere. However, the specific form of imbalance in convective-scale NWP is uncertain, and it is expected to differ from that in larger-scale models. To investigate this, three different imbalance measures each grounded in different physical rationales are implemented and applied to a convective resolving near-operational NWP system. This

includes a newly developed method based on a physical balance principle that applies approximately on the convective scale, the so-called “Weak Temperature Gradient (WTG) Balance”, which is being implemented and tested for the first time. The results show that the different measures seem to capture different facets of imbalance, with the WTG imbalance metric emerging as a particularly suitable candidate offering complementary results to existing methods.

In the future, we hope that the two presented diagnostic tools can contribute to further progress in the field of convective-scale DA, thereby enabling improvements in forecast accuracy.

Contents

Chapter 1

Introduction

> "Perhaps, some day in the dim future it will be possible to advance the computations faster than the weather advances ... But that is a dream"
>
> Lewis Fry Richardson, 1922

Approximately one hundred years ago, as of the time of writing this thesis, the scientist Lewis Fry Richardson envisioned a future where the weather could be predicted faster than it evolves (Richardson, 1922). Before articulating his vision, he embarked on the monumental task of manually computing the very first Numerical Weather Prediction (NWP). This endeavor involved discretising and integrating the equations of motion governing atmospheric dynamics. Although his prediction proved to be highly inaccurate, it should not be regarded as a failure. Given the limitations in "computational power" and the observational data he used as a starting point for his integration such an outcome was inevitable. Instead, it were Richardson's pioneering and brilliant ideas that served as the cornerstone for modern NWP models (Lynch, 2008).

The story of Richardson's first numerical weather prediction illustrates the achievements that have been made up to the present day. Over the past century, the seemingly insurmountable challenge of forecasting the future state of the atmosphere has transformed into a remarkable success story. In today's world, NWP has become an indispensable tool with high socio-economic value. Meteorological forecasts from operational centres are readily available all over the world.

Since the beginning of the 20th century NWP has undergone a "quiet revolution", as

Bauer et al. (2015) termed it, driven by the combination of scientific and technological progress. From a scientific point of view this progress has been mainly achieved by advancements in three realms of meteorology: 1. physical process representation, 2. ensemble forecasting, 3. model initialisation. Combined with technological progress, such as the availability of supercomputing power and advances in numerical algorithms for solving the equations of motions, the forecast skill has improved by about one forecast day per decade (Bauer et al., 2015).

The incorporation of a broad network of atmospheric observations makes a significant contribution to this success story, which is mostly important in the context of model initialisation. The integration of observational data into a numerical model to improve the system state estimation is called Data Assimilation (DA).

The research done in the frame of this dissertation lies in the field of DA. More specifically, it aims at providing efficient diagnostics, to better understand DA systems and therefore drive the development of improved algorithms and optimal use of observational data. In the long run this will yield improvements in forecasting skill (Gustafsson et al., 2018).

In the following, more detailed information and background knowledge on the state of the art of NWP and the specific challenges in convective-scale DA (Sections 1.1. and 1.2) will be given. The problem will be narrowed down to the specific research questions and objectives of this thesis in Section 1.3. Finally, Section 1.4 presents the outline of the thesis and Section 1.5 provides references to research articles published in the frame of this work.

1.1 Concepts of Numerical Weather Prediction

A contemporary NWP system typically consists of a forecasting model and a DA system. The forecasting model simulates the physical evolution of the atmosphere, which is governed by six prognostic equations: the Navier-Stokes equations (one for each component of the three dimensional wind velocity vector), the mass continuity equation, the first law of thermodynamics and the ideal gas law. These equations describe the spatial and temporal evolution of wind, pressure, density and temperature (see e.g. Vallis, 2017; Kalnay, 2002). Since there is no analytical solution for the set of prognostic equations they need to be discretised and integrated numerically, which constitutes an unavoidable approximation. The discretisation takes place on a three dimensional grid with finite resolution that spans the atmosphere (e.g. Zängl et al., 2015). Processes that are unresolved are usually referred to as subgrid-scale processes and their effect on the resolved grid-scale is approximated through parameterisations (Dudhia, 2014).

NWP is an initial value problem and the purpose of the DA system is to equip the forecasting model with initial conditions that represent the most accurate estimate of the true state of the atmosphere, considering all available information (see e.g. Daley, 1991; Talagrand, 1997). This best estimate of the atmospheric state is called the *analysis*. The information that is available to compute the analysis are atmospheric observations, however the observational coverage is much smaller than the number of degrees of freedom of the forecasting model and hence DA is a strongly under-determined problem. Thus, observational data is combined with prior information from a model forecast to obtain initial conditions that are defined on the entire numerical grid. This prior information is typically a short-range forecast initialised from a previous analysis, often referred to as *background* or *first-guess* forecast (Kalnay, 2002).

The analysis reflects the maximum likelihood estimate of the true state of the atmosphere, minimising the departures between the first guess and the observations, taking into account model and observation errors. This process, naturally, requires knowledge of, or assumptions on, model and observation errors, or more specifically their error covariances. Additionally, DA typically yields a new estimation of the error covariance associated with the analysis. In operational weather forecasting, different flavours of either a Kalman Filter or variational systems are used for DA. These algorithms both incorporate observational data at regular time intervals, though the specific approach to minimisation of the departures differs (see e.g. Potthast, 2019). More details on the Kalman Filter, which is the DA method employed in this work, can be found in Section 2.1.

In order to perform the minimisation to compute the analysis, model-generated prior information needs to be compared to observational data. Therefore, observation operators, also known as forward operators, are needed that map model variables from their state space into observation space. If the observed variable differs from a directly simulated prognostic variable, the observation operator may involve complex operations that rely on assumptions to facilitate the mapping process. For instance, assimilating visible satellite observations necessitates an observation operator that simulates cloud radiative transfer in order to translate the model prognostic variables into a model equivalent of observed cloud reflectance. This involves a variety of simplifying assumptions about the representation of clouds, and therefore contains errors (Scheck et al., 2016). The observation operator error is part of the observation error, along with contributions from two other error sources: the instrument error, and the error of representativeness, which arises from unresolved scales and processes within the numerical model; instruments measure point values or smaller cross-sections of the atmosphere, which may be influenced by atmospheric variability on spatial or temporal scales beyond the model's resolution. That means, even without instrument or operator error, observed values may differ from prior information, which must be taken into account (Janjić et al., 2018).

As mentioned earlier, both the background forecast and the analysis are subject to errors. It is crucial for DA, as well as for the overall utility of weather forecasting itself, to be able to assess this uncertainty. This is especially important due to the chaotic nature of our atmosphere, primarily attributed to the amplification of small inaccuracies in the initial conditions, which propagate across all temporal and spatial scales (Lorenz, 1963) and hence limit the predictability of weather.

In the present day, it is common practice to use ensembles to sample the uncertainty of the weather forecast. Ensemble forecasting involves the generation of multiple forecasting model runs with slight variations in initial conditions or model configurations. This approach provides probabilistic information about simulated weather conditions, accounting for the inherent uncertainties (Palmer, 2017). Due to the flow-dependency of uncertainties purely statistical approaches to quantify the uncertainty of the forecast are not sufficient. Multiple different fully non-linear realisations are crucial. The determination of initial perturbations that best represent the uncertainty of the system is a challenging task and a very active area of research (Buizza, 2019; Bauer et al., 2015). Variations in the initial conditions can be obtained through different techniques, including perturbing initial conditions or model parameters, providing various ways to span the ensemble. In modern NWP systems, ensemble predictions and DA are closely intertwined.

In Ensemble DA, a first guess ensemble is used to estimate the model error covariances, by correlating the perturbations from the ensemble mean. Since the first guess ensemble is produced from multiple model runs the equations of motion influence the way in which ensemble perturbations are correlated with each other, thereby allowing not only spatial correlations within one variable but also multivariate correlation between different prognostic variables (Bannister, 2008). Hence, the model error covariance matrix also reflects physical relationships which ideally should be preserved in the analysis.

Given the disparity between the available observational coverage and the model's spatial and temporal resolution, the error covariance matrix operates in a manner that effectively distributes the information gleaned from the observations (Kalnay, 2002). The job of Ensemble DA is to produce initial conditions for a whole analysis ensemble reflecting the flow-dependent uncertainty estimated in the analysis step (Evensen, 1994).

At the time of writing this thesis, weather services and research centers employ cycled NWP systems, which describes the alternation between weather forecasting and intermittent DA. Typical cycling intervals are on the order 1 to 6 hours, depending on the model resolution (Gustafsson et al., 2018). Horizontal resolutions of global models are of the order of 10 km, hence they are capable of resolving atmospheric phenomena such as fronts and cyclones, which are referred to as synoptic-scale weather phenomena (see e.g. Schraff et al., 2016; MetOffice, 2023). Since the early 2000s, NWP systems with grid-scale resolutions between 1 - 4 km are also used operationally and for research purposes (Gustafsson et al., 2018). NWP systems with such resolution are commonly referred to as convective-scale models since they are capable of resolving atmospheric phenomena, such as heavy precipitation produced by convective systems. Due to limited computational power, however, these models are not run globally, rather they are used to simulate the weather over a regional area, hence these models are also often referred to as limited area models. It is necessary to choose appropriate boundary conditions, as weather phenomena might enter or leave the simulated region. Typically, these are taken from global weather models which run at lower resolution and are then interpolated to the desired resolution of the regional model.

With the emergence of convective-scale ensemble simulations also the requirements on the DA have changed. Despite the significant advancements achieved in this field, forecasting convective-scale weather phenomena, known for their rapid changes and localised effects, continues to pose substantial challenges.

1.2 Challenges in Convective-Scale Data Assimilation

Although weather forecasting has steadily improved over the past century, there is still room for advancement. In particular, a more accurate estimation of the initial state could lead to more accurate weather forecasts at longer forecast times (Selz et al., 2022). In this context, also Bauer et al. (2015) identify convective-scale DA as one of the parts of NWP that bear the greatest challenges for future progress. Figure 1.1 (adopted from

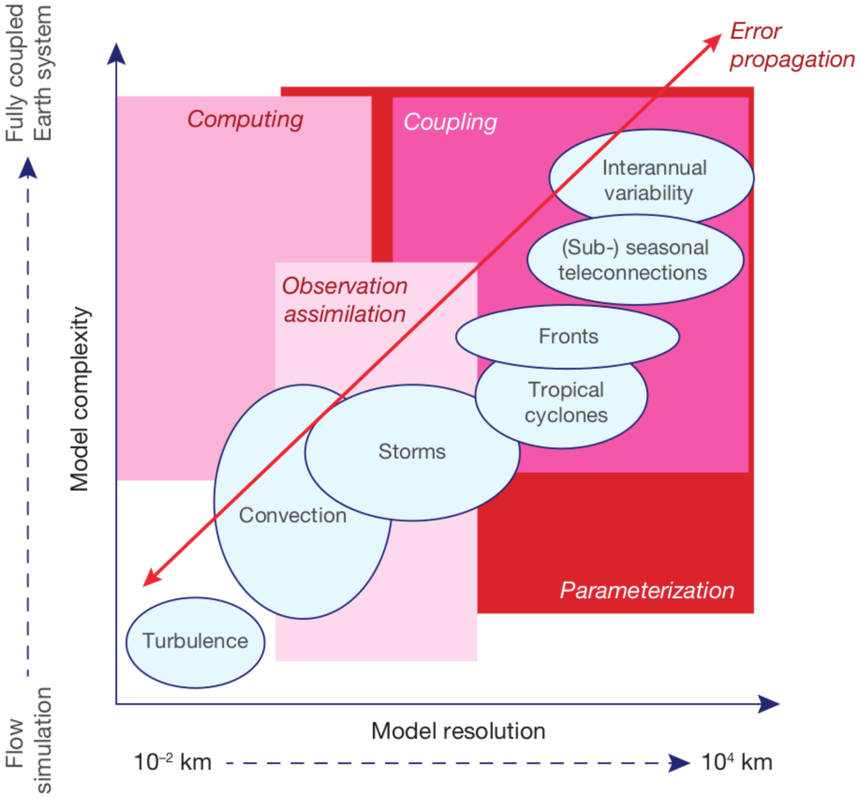


Figure 1.1: Figure adopted from Bauer et al. (2015). Key challenge areas for NWP in the future. The x-axis shows the model resolution, the y-axis indicates the model complexity. The boxes in different shades of red show the four main challenges in NWP. The blue elipses show different atmospheric phenomena aligned according to their prevailing scales and complexity. The red arrow indicates error propagation across the scales.

Bauer et al., 2015) provides an overview of key challenge areas for NWP, plotting resolved scales against model complexity on the x-axis and y-axis, respectively. The squared boxes indicate the main challenges for NWP, with convective-resolving models posing significant difficulties in the context of DA.

This arises from the fact that especially on the convective-scale, the nature of processes such as strong precipitation and thunderstorm is very nonlinear. Consequently, also initial condition errors grow much more rapidly on these scales. This limits the predictability of such processes to only a few hours (Hohenegger and Schär, 2007). Hence, there is need for frequent DA updates, which also necessitate highly resolved observational data. At the time of writing this thesis, however, there is little understanding of which update frequencies, variables and spatial scales are most efficient (Gustafsson et al., 2018; Lange and Craig, 2014; Greybush et al., 2011) for convective-scale DA. This is related to the fact that, while DA in principle provides the most accurate estimate of the initial state that is possible with the available information, it still contains a number of assumptions, approximations and tuning parameters that can limit the effectiveness of DA. Critical assumptions in modern DA systems used by operational centres include, for example, the Gaussianity of forecast and observation errors, linearity of observation operators, or the sampling of forecast uncertainty by the ensemble.

There are many different approaches to improving current methods. Efforts include the development of DA algorithms that do not rely on limiting assumptions such as Gaussianity, refinement of the observation network, or optimisation of DA parameters, which are the subject of ongoing research (Gustafsson et al., 2018). For all these research efforts, it is crucial to co-develop tools to evaluate the effectiveness of the DA system, which is often a methodological challenge in itself.

This thesis concentrates on the development of evaluation tools to assess the influence of observations on the analysis and in the short-term forecast. While it is particularly important to assess the influence of new, high-resolution observational data on the analysis when incorporating them into the DA system, the focus on the impact of observations in the short-term forecast is mainly on measuring whether the DA update generated by the observations provides initial states that are consistent with the model dynamics. This consistency is related to the physical error correlations within the model state and is often referred to as *balance*. These two aspects are explained in more detail in the following two subsections.

1.2.1 The Influence of Observations

Observation types such as geostationary satellites, radar, and ground-based remote sensing are a promising source of information for convective-scale DA because they deliver highly resolved information. However, their assimilation is challenging due to the complexity involved in creating observation operators and managing observation uncertainties (Hu et al., 2022). In fact, the assimilation of such observations became operational only in the recent years, even though they are still assimilated in systems which by design assume linear observation operators and in some cases uncorrelated observation errors, which is both violated in the case of radar and visible satellite images, for example. Before including these observations in the DA system it had to be ensured that their influence is beneficial for the weather forecast (see e.g. Scheck et al., 2020; Geiss et al., 2021) through the use of observation influence diagnostics. Operational weather forecasting centers also utilize different kinds of diagnostic tools to assess the effect of observations and to continuously monitor their DA systems (Casaretto et al., 2023).

These diagnostic tools can be categorised based on the time frame being assessed. “Observation influence diagnostics” focus on the sensitivity of analysis to observations, while “observation impact diagnostics” assess the value of observations at specific forecast lead times. The methods for computing observation influence or impact are further distinguished by direct and indirect approaches (Hu et al., 2023).

The first example of a direct method are Observing System Experiments (OSE, Bouttier and Kelly, 2001), also known as data denial experiments. In an OSE, all available data are assimilated into an NWP model, followed by a second experiment in which certain data are intentionally excluded. The difference between the results of these two experiments indicates the influence of the omitted data on the performance of the model (Eyre, 2021). A second type of direct experiments are Observation System Simulation Experiment (OSSE), which are similar to OSE, except that the observing system is simulated from a “nature run”, which represents a modeled atmospheric state, which is assumed to be the truth. Synthetic observations are drawn from the nature run using observation operators and adding a random error to it. The benefit of an OSSE over an OSE is firstly that researchers have more control over the DA experiment and secondly an OSSE allows for simulating the potential impact of novel and experimental observation technologies (see e.g. Errico et al., 2013). Both OSE and OSSE require conducting at least two DA experiments to assess the effects of different observational datasets on model performance.

The second class of methods are indirect methods based on Forecast Sensitivity to Observations (FSO, Langland and Baker, 2004). This means, they measure the extent to which a subset of observations contributes to the reduction of analysis or short-range forecast error within a system assimilating all observations. Kalnay et al. (2012) demonstrated how to compute Ensemble Forecast Sensitivity to Observations (EFSO), measuring the impact of observations within an Ensemble DA system.

While data denial experiments may represent a more accurate method for assessing impact and influence, they are computationally expensive and the best strategy for assessing the value of thousands of different assimilated observations through data denial is unknown. EFSO methods allow for a computationally more efficient assessment of the effect of the assimilated observation and are therefore essential on the convective scale. While studies such as Sommer and Weissmann (2016) or Necker et al. (2018) have successfully applied EFSO in convective-scale DA systems, its computation relies on the assumption of a linear forecasting model and therefore delivers only approximate results (see Kalnay et al., 2012; Sommer and Weissmann, 2014). For these reasons, recent reviews in the field, such as Gustafsson et al. (2018) and Hu et al. (2023), suggest that supplementary observation influence and impact diagnostics may help to guide monitoring, planning and decision making on the integration of novel observation types and future observing systems.

1.2.2 Localisation and Imbalance

A key factor for efficient DA is that the produced analysis has to be consistent with the dynamics of the forecasting model, since the analysis serves as the initial condition for the subsequent forecast. This form of consistency can be imagined as a balance of forces that govern the dynamical evolution of the atmosphere. The true observed motion of the atmosphere appears to be confined to a subset of possible motions, that can be characterised by predominant balances of forces. This subset of possible motions forms an attractor, in the sense that deviations from balance decay rapidly in comparison to the evolution of the balanced motions. In DA, the computed analysis is not necessarily close to the attractor, hence it is *imbalanced*, and the model will adjust towards a balanced state as soon as the new forecast simulation starts. This can result in unrealistic fields early in the forecast as the adjustment process proceeds, and can lead to spurious modifications of the balanced part of the flow that reduce skill at longer lead times. A famous example of how noisy data and an imbalanced initial state can lead to big mistakes was the first numerical weather prediction of Richardson in 1922, mentioned at the beginning of this work. Due to spurious inertia-gravity waves, he predicted a change in surface pressure of 146 hPa within 6 hours, whereas, in reality, the surface pressure stayed almost unchanged (Kalnay, 2002). The reason for imbalanced analyses are inaccuracies in the estimation of the model error covariance matrix. As mentioned above, the model error covariance matrix is responsible for distributing the observational information within the model domain according to its inherent physical correlations. If these correlations are incorrectly specified, the resulting DA update may violate physical relationships.

In Ensemble DA, a critical approximation involves sampling the uncertainty of background errors using an ensemble. Since ensembles come with a high computational cost, typical ensemble sizes in regional and global NWP are typically limited to between 10 and 50 members (see e.g. MetOffice, 2023; Reinert et al., 2020; Buizza, 2019). The limited number of ensemble members typically leads to undersampling of the real model error covariance and may introduce spurious correlations, which are erroneous or coincidental relationships between variables in the forecast. These misleading correlations can, in turn, impact forecasting skill by introducing noise or false signals into the prediction (Lange and Craig, 2014).

To mitigate the influence of spurious correlations caused by limited ensemble size, localisation techniques are employed in ensemble forecasting. Localisation involves limiting the influence of observations to a specified spatial domain, helping to filter out spurious long-distant correlations that are often associated with noise. This approach aids in reducing the impact of spurious correlations by focusing on locally coherent patterns and relationships within the ensemble (Evensen et al., 2022).

The perfect localisation length scale (if that even exists), however, is unknown and may be different depending on the considered variables and weather situation. Typically, the localisation length scale is chosen such that the number of observations that influence a specific grid point is on the order of the ensemble size. The rationale behind that is that the number of degrees of freedom of the model within a characteristic localisation radius is roughly determined by the ensemble size (Schraff et al., 2016). However, localisation still represents an ad-hoc method and its parameters are subject to tuning. Consequently, it can happen that localisation will cut off covariances that are actually physically correct. This implies that ensemble DA in combination with localisation ensures consistency between the analysis state and the observations locally, but it may result in inconsistencies with respect to the model dynamics.

Due to the potential negative consequences of localisation and associated imbalances it is important to monitor and counteract imbalance in DA systems.To reduce imbalance in initial conditions, NWP systems use a variety of techniques, often referred to as initialisation. For the synoptic scale, the notion of imbalance is associated with the violation of predominant balances of forces, such as hydrostatic and geostrophic balance. Due to these balance principles there is a clear time scale separation between fast transient motion and slow balanced flow. Since geostrophic motions are non-divergent, a simple method of reducing ageostrophic motions is to damp the divergent wind. A more sophisticated method is Nonlinear Normal Mode Initialisation (Tribbia, 2020), where the flow is decomposed into a spectrum of modes and the tendencies of the fast modes are set to zero, producing an initial state that will evolve mainly on slow time scales. A pragmatic way of reducing the effects of fast modes is to apply a digital filter to damp fast variability early in the simulation (Lynch and Huang, 1992). Finally, a balance principle such as geostrophic balance or a generalisation like linear balance can be introduced by penalising departures from balance as part of the DA optimisation. Methods such as these, singly or in combination, are a standard part of modern NWP systems (Berre, 2000; Peckham et al., 2016).

On the convective-scale, however, the situation is less clear. Geostrophic and hydrostatic balance are no longer good assumptions on these scales (Vetra-Carvalho et al., 2012) and hence initialisation techniques need to be adapted. While there has been considerable research in recent years aimed at addressing the problem of imbalance resulting from DA at the convective scale (Gustafsson et al., 2018), relatively little attention has been paid to developing more appropriate metrics for measuring this imbalance (e.g., Zeng et al., 2021; Ge et al., 2022). More appropriate ways to measure imbalance on the convective-scale might help identify potential flaws in the DA system, quantify the effectiveness of assimilating certain observation types in general, or the effectiveness of initialisation techniques.

Summarising this, for convective-scale DA we find the following situation: There is need for frequent DA updates, using complex atmospheric observations, which may violate assumptions in the DA system. Therefore, the assessment of the value of observations in the DA system is crucial, but complex or potentially computationally intractable. Moreover the dynamics on the convective scale are highly non-linear, known balance principles may not be appropriate for the convective scale and there is lack of systematic research on ways to measure or counteract imbalance introduced through convective-scale DA.

1.3 Objectives and Research Questions of this Thesis

Modern atmospheric observations have significant potential for improving short-term weather forecasts. The assessment of observation influence statistics are an important component in this development. However, establishing a meaningful connection between currently used observation influence statistics and the physical evolution of the atmosphere is challenging. The overarching goal of this research project is to address the two key challenges identified in the preceding discussion, namely to develop novel methodologies to quantify analysis influence of observations and to measure imbalances caused by data assimilation. Therefore this work adresses the following research questions:

Research Question 1:	How can we evaluate the impact of various observations in a way that yields easily interpretable results and ensures computational efficiency?
Research Question 2:	Which measures of imbalance are suitable for the convective scale?

1.4 Outline

This thesis is structured as follows: Chapter 2 provides the basic principles and background knowledge forming the basis on which this research is built upon. It describes the theory of the Local Ensemble Transform Kalman Filter (Section 2.1) and the current understanding of imbalance introduced through data assimilation (Section 2.2)

The main results of our research are presented in Chapters 4 and 5. Chapter 4 deals with the partial analysis increments diagnostic and Research Question 1 is considered in detail. Chapter 5 discusses the performance of imbalance metrics suitable for the convective scale. It forms the second major research segment of this work, focusing on Research Question 2.

The thesis is concluded in Chapter 6 by summarising and discussing the principal findings of this work in light of the aforementioned research objectives. Finally, this chapter also outlines possible future avenues of research aimed at further enhancing our understanding of the influence of observations and imbalance in convective-scale data assimilation.

1.5 Publications

Parts of this dissertation are published as the following publications:

(a) Diefenbach, T., Craig, G., Keil, C., Scheck, L. & Weissmann, M.(2023) Partial analysis increments as diagnostic for LETKF data assimilation systems. Quarterly Journal of the Royal Meteorological Society, 149(752), 740–756. Available from: `https://doi.org/10.1002/qj.4419`

(b) Diefenbach, T., Scheck, L. & Weissmann, M., Craig, G.(2024) Diagnostics for Imbalance on the Convective Scale, submitted to Monthly Weather Review

Chapter 2

Theory

In this thesis we investigate the influence of observations on the analysis and imbalance that is introduced through data assimilation. The goal of this chapter is to review the theory and concepts that are applied in the context of this work. Section 2.1 reviews the theory and the mathematical framework of data assimilation. In particular, it provides an explanation of the Local Ensemble Transform Kalman Filter, which is the method used in this work. Section 2.2 lays a theoretical foundation to understand imbalance, both from a dynamical point of view, as well as from a data assimilation perspective.

2.1 The Local Ensemble Transform Kalman Filter

The Kalman filter is a sequential data assimilation method that optimally estimates the evolving state of a dynamic system by combining new observations with prior state estimates. The Local Ensemble Transform Kalman Filter (LETKF) is a form of the Kalman Filter which is, by design, suited for the needs of high-dimensional systems and large numbers of observations. This chapter derives the LETKF equations step by step. We start with important definitions and introduce the concept of the Kalman Filter. Then, we describe the adaptations made in the Ensemble Kalman Filter and introduce the Ensemble Transform Kalman Filter. Finally, localisation is explained. Except where noted, the derivations given in this section follow Kalnay (2002) and Hunt et al. (2007). The notation follows Hunt et al. (2007).

2.1.1 Definition of the Data Assimilation Problem and the Kalman Filter

The goal of data assimilation is to estimate the true state of the atmosphere $\boldsymbol{x}^t$ as accurately as possible. The best estimate of the state of the atmosphere is called the analysis, denoted by $\boldsymbol{x}^a$. The information that is available is data from observations as well as from previous model forecasts. We denote the previously modeled state as $\boldsymbol{x}^b$ and the observations as $\boldsymbol{y}^o$. Note that since the coverage of observations is generally much smaller than the degrees of freedom of the forecasting model, the previous forecasts are essential to distribute the information from past observations over space and time. $\boldsymbol{x}^t$, $\boldsymbol{x}^a$ and $\boldsymbol{x}^b$ are represented as m-dimensional vectors, where m is the number of model variables. $\boldsymbol{y}^o$ is a vector of size p, where p is the number of available observations. A weather forecasting model aims to simulate the true state of the atmosphere based on the laws of physics that govern it. The model predicted state is commonly referred to as *first guess* or *background* forecast (that is where the index b comes from). The relation between $\boldsymbol{x}^t$ and $\boldsymbol{x}^b$ can be written as

$$\boldsymbol{x}^b = \boldsymbol{x}^t + \boldsymbol{\varepsilon}^b, \tag{2.1}$$

where $\boldsymbol{\varepsilon}^b$ is the model error vector. The relation between observations and the true system state can be established as

$$\boldsymbol{y}^o = H(\boldsymbol{x}^t) + \boldsymbol{\varepsilon}^o \tag{2.2}$$

with $H(\boldsymbol{x}^t)$ being the observation operator that transforms from m-dimensional model space into p-dimensional observation space. $\boldsymbol{\varepsilon}^o$ is the observation error vector. Similarly, for the analysis we can write

$$\boldsymbol{x}^a = \boldsymbol{x}^t + \boldsymbol{\varepsilon}^a, \tag{2.3}$$

with $\boldsymbol{\varepsilon}^a$ being the analysis error vector.

The main assumptions that are made are that the errors are Gaussian and unbiased, with

$$\mathbb{E}(\boldsymbol{\varepsilon}^b) = \mathbb{E}(\boldsymbol{\varepsilon}^a) = \mathbb{E}(\boldsymbol{\varepsilon}^o) = 0 \tag{2.4}$$

Further, it is assumed that the model and observation errors are uncorrelated,

$$\mathbb{E}(\boldsymbol{\varepsilon}^b \boldsymbol{\varepsilon}^{oT}) = 0. \tag{2.5}$$

The model, analysis and observation error covariances are defined as

$$\mathbf{P}^b = \underbrace{\mathbb{E}(\boldsymbol{\varepsilon}^b \boldsymbol{\varepsilon}^{bT})}_{[m\times m]}, \quad \mathbf{P}^a = \underbrace{\mathbb{E}(\boldsymbol{\varepsilon}^a \boldsymbol{\varepsilon}^{aT})}_{[m\times m]}, \quad \mathbf{R} = \underbrace{\mathbb{E}(\boldsymbol{\varepsilon}^o \boldsymbol{\varepsilon}^{oT}).}_{[p\times p]} \tag{2.6}$$

To compute the analysis at any given time t_n the Kalman Filter seeks to minimise a cost function of the form

$$\mathcal{J}_{t_n}^{KF}(\boldsymbol{x}) = \frac{1}{2} \underbrace{(\boldsymbol{x} - \boldsymbol{x}_n^b)^T (\mathbf{P}_n^b)^{-1} (\boldsymbol{x} - \boldsymbol{x}_n^b)}_{\mathcal{J}_{tn,b}^{KF}} + \frac{1}{2} \underbrace{[\boldsymbol{y}_n^o - H_n(\boldsymbol{x})]^T \mathbf{R}_n^{-1} [\boldsymbol{y}_n^o - H_n(\boldsymbol{x})]}_{\mathcal{J}_{tn,o}^{KF}} . \tag{2.7}$$

The cost function is designed such that the analysis fits the observations and background best, in a least squares sense, which also reflects the maximum likelihood estimate of measuring $\boldsymbol{x}$ given $\boldsymbol{x}^b$ and $\boldsymbol{y}^o$. In the term $\mathcal{J}_{tn,b}^{KF}$ the information of all previous observations up to time t_{n-1} is gathered, in the sense that an analysis was computed at time t_{n-1} and then propagated in time using the forecasting model. The Kalman Filter assumes a linear model operator, denoted as $\mathbf{M}_{t_{n-1},t_n}$ that propagates the system state forward in time,

$$\boldsymbol{x}_n^b = \mathbf{M}_{t_{n-1},t_n} \boldsymbol{x}_{n-1}^a, \tag{2.8}$$

$$\mathbf{P}_n^b = \mathbf{M}_{t_{n-1},t_n} \mathbf{P}_{n-1}^a \mathbf{M}_{t_{n-1},t_n}^T, \tag{2.9}$$

The term $\mathcal{J}_{tn,o}^{KF}$ adds the information coming from the current set of observations. This results in an iterative algorithm, with alternating analysis and forecast steps. The process is pictured in the schematic Figure 2.1.

To solve the cost function analytically, linearisation of $H(\boldsymbol{x})$ is performed through a Taylor expansion around $\boldsymbol{x}^b$. That means

$$H(\boldsymbol{x}) \approx H(\boldsymbol{x}^b) + \mathbf{H}|_{\boldsymbol{x}=\boldsymbol{x}^b} (\boldsymbol{x} - \boldsymbol{x}^b) \tag{2.10}$$

with $\mathbf{H}|_{\boldsymbol{x}=\boldsymbol{x}^b}$ being the Jacobian of H with elements $h_{(ij)} = \frac{\partial h_{(i)}}{\partial x_{(j)}}$ evaluated at $\boldsymbol{x} = \boldsymbol{x}^b$. To simplify the notation we will simply write $\mathbf{H}$, meaning $\mathbf{H}|_{\boldsymbol{x}=\boldsymbol{x}^b}$, in the proceeding.

The condition for finding the minimum of $\mathcal{J}_{t_n}^{KF}$ is

$$\nabla_{\boldsymbol{x}} \, \mathcal{J}_{t_n}^{KF}(\boldsymbol{x}) \overset{!}{=} 0. \tag{2.11}$$

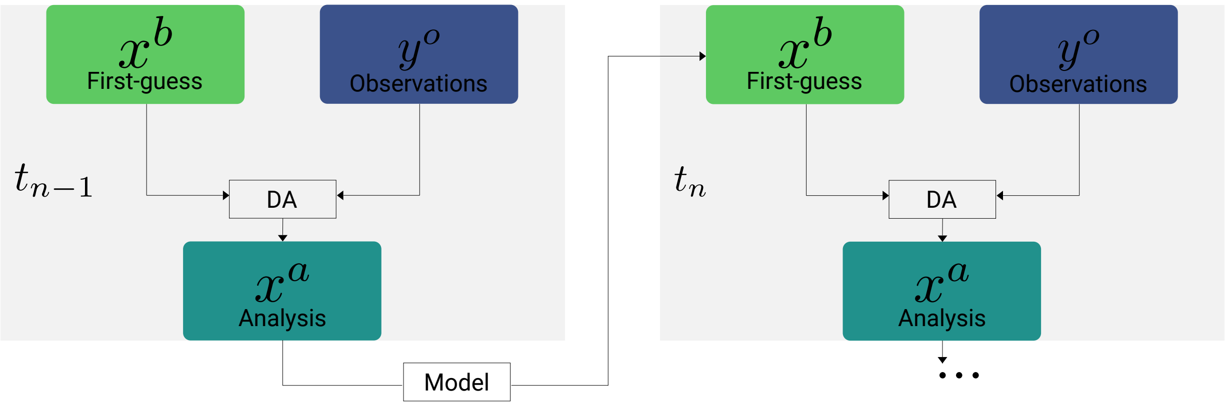


Figure 2.1: Schematic diagram of two data assimilation cycles at t_{n-1} and t_n.

Using Eq. (2.10) in Eq. (2.7) yields a cost function that is quadratic in $\boldsymbol{x}$ and hence

$$\nabla_{x}\mathcal{J}_{t_n}^{KF} = \mathbf{H}_n^T\mathbf{R}_n^{-1}\left[H_n(\boldsymbol{x}_n^b) + \mathbf{H}_n(\boldsymbol{x}-\boldsymbol{x}_n^b) - \boldsymbol{y}_n^o\right] + (\mathbf{P}_n^b)^{-1}(\boldsymbol{x}-\boldsymbol{x}_n^b) \overset{!}{=} 0. \tag{2.12}$$

Solving for $\boldsymbol{x}$ yields the analysis equation,

$$\boldsymbol{x}_n^a = \boldsymbol{x}_n^b + \underbrace{\underbrace{[(\mathbf{P}_n^b)^{-1} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n]^{-1}}_{\mathbf{P}_n^a,[m\times m]}\mathbf{H}_n^T\mathbf{R}_n^{-1}}_{\mathbf{K}_n,[m\times p]}\left[\boldsymbol{y}_n^o - H_n(\boldsymbol{x}_n^b)\right]. \tag{2.13}$$

The analysis covariance at time t_n can then be computed as

$$\mathbf{P}_n^a = \mathbb{E}[(\boldsymbol{x}_n^a - \boldsymbol{x}_n^t)^2] = [(\mathbf{P}_n^b)^{-1} + \mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n]^{-1}. \tag{2.14}$$

For convenience, we will drop the time indices n in the proceeding, however all variables will still refer to only one time step t_n. The term $[(\mathbf{P}^b)^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{K}$ is the weighting matrix, often referred to as the *Kalman Gain*. With this definition the sequential update equation of the Kalman Filter becomes

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \mathbf{K}\left[\boldsymbol{y}^o - H(\boldsymbol{x}^b)\right]. \tag{2.15}$$

The term $\mathbf{K}\left[\boldsymbol{y}^o - H(\boldsymbol{x}^b)\right]$ is called the *increment*. In this form it becomes obvious that the analysis is a weighted sum of the background and the observations and that knowledge of $\mathbf{R}$ and $\mathbf{P}^b$ are essential to determine the weighting.

As already mentioned in the introduction the observation error and, with this, its covariance matrix $\mathbf{R}$ can be estimated using e.g. the Desroziers method (Desroziers et al., 2005). In the proceeding of this work, $\mathbf{R}$ will be generally assumed as given and as a diagonal matrix, i.e. the errors of individual observations are uncorrelated.

Further, we want to point out the crucial role of $\mathbf{P}^b$. Besides the error variances in the diagonal, $\mathbf{P}^b$ inheres autocorrelations between the same variables at different grid points. Further $\mathbf{P}^b$ also inheres multivariate correlations among different variables. This means, $\mathbf{P}^b$ is responsible for spreading observational information vertically and horizontally in space and among different variables, which should ensure physical balance. Therefore, accurate representation of $\mathbf{P}^b$ is essential. While various methods exist to model $\mathbf{P}^b$, its typically very large matrix size renders pure Kalman filtering impractical for large systems due to the associated high computational costs of multiplying such matrices. One way to model $\mathbf{P}^b$ is through ensembles, which will be explained in the following subsection.

So far, we have not commented on how to initialise $\mathbf{P}^b$ for the very first analysis in the iteration. Formally, an infinite $\mathbf{P}_1^b$ could be assumed. However, if the number of observations at t_1 is much smaller than the number of model variables, then $\mathbf{H}_1$ is not invertible and Eq. 2.14 cannot be solved. In practice, to avoid this difficulty a prior $\mathbf{P}_1^b$ is assumed which is reasonably large but finite. The effect of this assumption decreases as the number of assimilated observations increases over time.

2.1.2 The Ensemble Kalman Filter

The idea of the Ensemble Kalman Filter is to use an ensemble of system states to sample the probability density function of the atmospheric state (Evensen, 1994). Each ensemble member will be propagated in time individually using the nonlinear model operator $\mathcal{M}$. The ensemble mean and covariance is defined as

$$\overline{\boldsymbol{x}}^b = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{x}^b_{(i)}, \qquad \mathbf{P}^b = \frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{x}^b_{(i)} - \overline{\boldsymbol{x}}^b)^T(\boldsymbol{x}^b_{(i)} - \overline{\boldsymbol{x}}^b) = \frac{1}{k}\mathbf{X}^b\mathbf{X}^{bT} \tag{2.16}$$

$$\overline{\boldsymbol{x}}^a = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{x}^a_{(i)}, \qquad \mathbf{P}^a = \frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{x}^a_{(i)} - \overline{\boldsymbol{x}}^a)^T(\boldsymbol{x}^a_{(i)} - \overline{\boldsymbol{x}}^a) = \frac{1}{k}\mathbf{X}^a\mathbf{X}^{aT} \tag{2.17}$$

where $\mathbf{X}^v$, $v \in [a, b]$ are the ensemble perturbation matrices, for the background or the analysis, respectively; each column of $\mathbf{X}^v$ is the deviation of one ensemble member i and the ensemble mean $\overline{\boldsymbol{x}}^v$. The dimension of $\mathbf{X}^v$ is $m \times k$, where k is the number of

ensemble members.
The ensemble allows for a flow-dependent estimation of the model error covariance. However, the ensemble-estimated $\mathbf{P}^b$ has rank at most $k-1$ and is therefore rank deficient and not invertible. Additionally, the computation of the analysis would still involve the large $\mathbf{P}^b$ matrix. A more efficient approach to solving the Kalman Filter cost function is the Ensemble Transform Kalman Filter.

2.1.3 The Ensemble Transform Kalman Filter

The trick of the Ensemble Transform Kalman Filter (ETKF) is to transform the state variables into a space spanned by the ensemble. The minimisation is then carried out in the low-dimensional ensemble space instead of high-dimensional model space.
This can be done by regarding the ensemble perturbations as a linear transformation that transforms a k-dimensional vector $\boldsymbol{w}$ from ensemble space into model space. $\mathbf{X}^b\boldsymbol{w}$ is then a vector in the space spanned by the background ensemble perturbations and

$$\boldsymbol{x} = \overline{\boldsymbol{x}}^b + \mathbf{X}^b\boldsymbol{w} \tag{2.18}$$

is a new model state. Inserting this representation of $\boldsymbol{x}$ into the KF cost function ($\mathcal{J}^{KF}$, Eq. 2.7) yields the ETKF cost function in ensemble space:

$$\mathcal{J}^{ETKF}(\boldsymbol{w}) = (k-1)\boldsymbol{w}\boldsymbol{w}^T + (\boldsymbol{y}^o - H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b\boldsymbol{w}))^T\mathbf{R}^{-1}(\boldsymbol{y}^o - H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b\boldsymbol{w})). \tag{2.19}$$

If we find a $\overline{\boldsymbol{w}}_a$ that minimises $\mathcal{J}^{ETKF}$ then

$$\overline{\boldsymbol{x}}^a = \overline{\boldsymbol{x}}^b + \mathbf{X}^b\overline{\boldsymbol{w}}^a, \tag{2.20}$$

minimises the cost function $\mathcal{J}^{KF}$. This means the analysis ensemble mean $\overline{\boldsymbol{x}}^a$ is constructed using a linear combination of the background ensemble members with optimal weights $\overline{\boldsymbol{w}}^a$.
It would be possible to allow for a nonlinear observation operator and minimise $\mathcal{J}^{ETKF}$ numerically. However, for formulating the analysis more explicitly H can be linearised. The simplest way to linearise H is to apply it to the ensemble members individually and then interpolate. This means that

$$\boldsymbol{y}^b_{(i)} = H(\boldsymbol{x}^b_{(i)}). \tag{2.21}$$

Let us define $\mathbf{Y}_b = \boldsymbol{y}^b_{(i)} - \overline{\boldsymbol{y}}^b$ and hence

$$H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b \boldsymbol{w}) \approx \overline{\boldsymbol{y}}^b + \mathbf{Y}^b \boldsymbol{w}. \tag{2.22}$$

Note that additivity of the H function is assumed, but the H that is applied to the terms on the right hand side of the equation is still the non-linear observation operator. In other words this approximation means that the model equivalent of the analysis can be obtained by creating a linear combination of the first guess ensemble model equivalents. The linear approximation yields the cost function

$$\tilde{\mathcal{J}}^{ETKF}(\boldsymbol{w}) = (k-1)\boldsymbol{w}^T\boldsymbol{w} + (\boldsymbol{y}^o - \overline{\boldsymbol{y}}^b + \mathbf{Y}^b\boldsymbol{w})^T\mathbf{R}^{-1}(\boldsymbol{y}^o - \overline{\boldsymbol{y}}^b + \mathbf{Y}^b\boldsymbol{w}). \tag{2.23}$$

This quadratic equation can now be solved analytically. Leading to

$$\overline{\boldsymbol{w}}^a = \tilde{\mathbf{P}}^a \mathbf{Y}^{bT}\mathbf{R}^{-1}(\boldsymbol{y}^o - \boldsymbol{y}^b), \tag{2.24}$$

$$\tilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b]^{-1}, \tag{2.25}$$

with the back transformation into model space

$$\overline{\boldsymbol{x}}^a = \overline{\boldsymbol{x}}^b + \mathbf{X}^b \overline{\boldsymbol{w}}^a, \tag{2.26}$$

$$\mathbf{P}^a = \mathbf{X}^b \tilde{\mathbf{P}}^a \mathbf{X}^{bT}. \tag{2.27}$$

With this, $\overline{\boldsymbol{x}}^a$ and $\mathbf{P}^a$ are explicitly specified.

The next step is to resample the full analysis ensemble with the computed $\overline{\boldsymbol{x}}^a$ and $\mathbf{P}^a$. The ETKF following Hunt et al. (2007) uses the symmetric square root to span the ensemble,

$$\mathbf{W}^a = \sqrt{(k-1)\tilde{\mathbf{P}}^a} \tag{2.28}$$

$$\Leftrightarrow \qquad \tilde{\mathbf{P}}_a = \frac{1}{k-1}\mathbf{W}^a\mathbf{W}^{aT}, \tag{2.29}$$

with $\mathbf{W}^a$ reflecting the ensemble perturbations in ensemble space. $\mathbf{W}^a$ has dimension $k \times k$; each column i of $\mathbf{W}_a$ contains a weight vector $\boldsymbol{w}^a_{(i)} - \overline{\boldsymbol{w}}^a$ that determines the weights in the linear combination of the background ensemble members that are used to form an individual ensemble member i. With this, the analysis of one ensemble member in model space is given by

$$\boldsymbol{x}^a_{(i)} = \overline{\boldsymbol{x}}^b + \mathbf{X}^b \boldsymbol{w}^a_{(i)}. \tag{2.30}$$

Using the symmetric square root function to determine the ensemble weight vectors in ensemble space is a choice that has several advantages. Firstly, it ensures that the columns of the resulting analysis ensemble perturbations matrix $\mathbf{X}^a$ add up to 0, which ensures that the analysis ensemble has the correct ensemble mean. Secondly, $\mathbf{W}^a$ depends continuously on $\tilde{\mathbf{P}}^a$. This is especially important in the context of localisation, which will be explained in more detail in the next subsection.

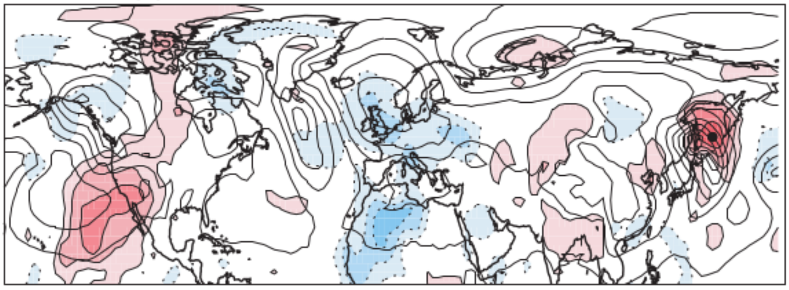
2.1.4 Localisation

While the ETKF is advantageous from the point of view of computational efficiency it has two disadvantages: 1. It involves an approximation, since the analysis ensemble mean is restricted to the space spanned by the background ensemble. 2. The ensemble error covariance matrix is subject to sampling error, which may manifest itself through spurious correlations within the ensemble. Both effects can be reduced by increasing the ensemble size. For the sampling error this is illustrated in Figure 2.2 which is taken from Hamill (2006). In Panels a and b the color shows the correlation in $\mathbf{P}^b$ of sea level pressure with pressure at a point in the western pacific, indicated by the black dot. The solid black lines indicate the ensemble mean. In Panel a the correlations are estimated directly from a 25-member ensemble. In Panel b the correlations are estimated from a 200-member ensemble. Since sampling error decreases with increasing ensemble size the additional correlations estimated by the 25-member ensemble can be regarded as artefacts of limited ensemble size. Since a higher number of ensemble members is often not affordable localisation can be used in order to get rid of undesired correlations.

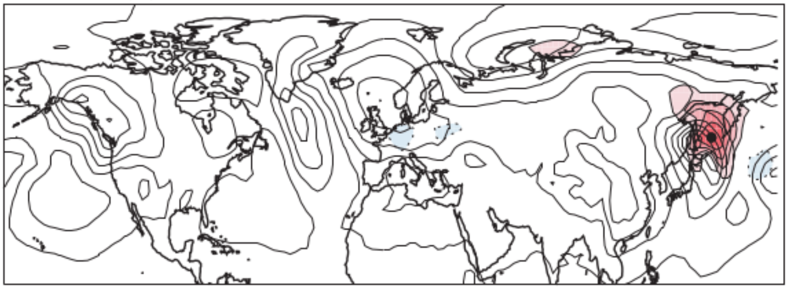
The most obvious approach for localisation is to set longer distance correlations in the model error covariance matrix to zero, as they are most likely spurious. This is not done in the LETKF implementation used in this work. One reason for that is because the whole background error covariance matrix is never used explicitly in the LETKF.

Instead, the influence of the observations is limited in space through localisation of $\mathbf{R}$. While this approach does not yield the exact same result as $\mathbf{P}^b$ localisation, the desired effect of cutting off long distance correlations stays the same. Mathematically, $\mathbf{R}$-localisation works through gradually increasing the observation error until, beyond a certain distance, the error is infinite and the influence on the analysis is zero. In prac-

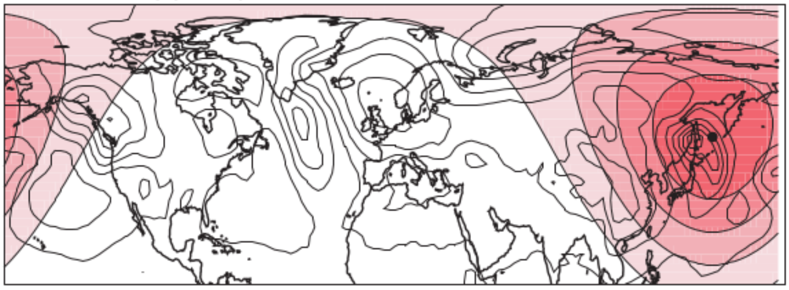
(a) Correlations in P^b, 25−member ensemble



(b) Correlations in P^b, 200−member ensemble



(c) Gaspari and Cohn correlation function



(d) Correlations in P^b after localisation, 25−member ensemble

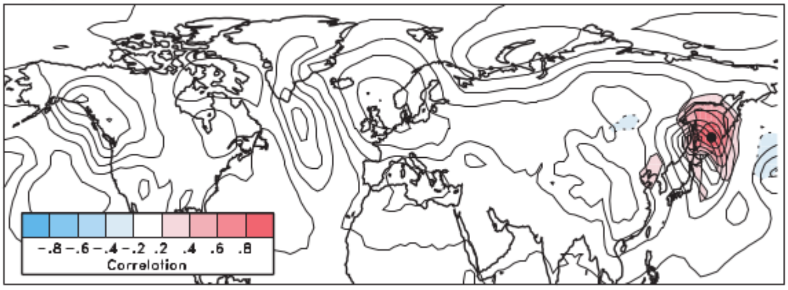


Figure 2.2: Illustration of covariance localisation. (a) Correlations of sea-level pressure directly estimated from 25-member ensemble with pressure at a point in the western Pacific (colors). Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa. (b) As in (a), but using 200-member ensemble. (c) Covariance localisation correlation function. (d) Correlation estimate from 25-member ensemble after application of covariance localisation. This figure and caption are taken from Hamill (2006)

tice, $\mathbf{R}$ is multiplied by a distance-dependent Gaussian shaped function that equals one at the nominal position of the observation and goes to zero beyond a certain distance. This function was introduced by Gaspari and Cohn (1999) and is therefore referred to as the Gaspari-Cohn function. A two-dimensional, horizontal Gaspari-Cohn function is depicted in Figure 2.2, Panel c, where the color indicates the value of the Gaspari-Cohn function ranging from zero to one. Figure 2.2, Panel d, shows the sea level pressure correlations as estimated by the 25-member ensemble after localisation with the Gaspari-Cohn function. It is evident that the spurious correlations have been eliminated, resulting in a correlation pattern more similar to that estimated by the 200-member ensemble. Hence, localisation of the 25-member estimate has a positive effect, since it reduces the artefacts due to undersampling of the true error covariance.

Due to the need for localisation, a local analysis is computed for each grid point in the analysis grid. This means that every analysis grid point “sees” a different subset of observations, and hence the minimisation of the LETKF cost function is performed individually at each location. Here, the smoothness of the square-root filter comes into play since it ensures, that also for a slightly different $\tilde{\mathbf{P}}_a$ at nearby grid points the constructed analysis ensemble varies smoothly. A local ETKF implementation has other advantages such as the effective increase in the degrees of freedom of the system, since every analysis grid point is considered independently of every other. That also means that the minimisation can be computed in parallel which reduces the computing time and makes the method feasible. While localisation is crucial to avoid sampling noise, it comes at the cost of potentially cutting off physically correct correlations, which has implications for the physical balance of the system, which will be discussed in Section 2.2.

2.1.5 Concluding Remarks on Observation Influence

In DA the Kalman Gain matrix determines the influence of the observations on the analysis and the weight that is given to the background forecast. Especially, when it comes to the design and evaluation of an LETKF system and the associated observational network, determination of the observation influence is important. However, as we have seen, in the LETKF the Kalman Gain matrix is never computed explicitly. Moreover, localisation implies that an independent set of analysis ensemble weights is computed at every analysis grid point. This means, the information provided by the observations is essentially cached in the local ensemble weights, and it is a challenging task to disentangle this information and relate it back to the contribution of individual observations to the analysis. Chapter 3 presents a method to compute the analysis influence of observations in an LETKF system in a computationally efficient way, and its potential applications.

2.2 Imbalance

So far in this chapter, we have explained how the local ensemble transform filter incorporates observational information into the model state. However, this incorporation can lead to a dynamically inconsistent state. Initialisation from a dynamically inconsistent state triggers adjustment processes that can limit the effectiveness of data assimilation and even degrade the forecast. In the following subsections we will explain the notion of imbalance from different perspectives. Section 2.2.1 will explain the concept of imbalance (and balance) from a dynamical point of view. In Section 2.2.2, we will address the issue in the context of data assimilation.

It is worth noting that, in the context of convective-resolving NWP, the concept of imbalance is elusive and lacks a precise and commonly agreed definition. At the time of writing this thesis it remains an open question if a clear definition even exists. Consequently, this section attempts to consolidate the current state of knowledge on the subject.

2.2.1 Balance Principles

The atmosphere is subject to multiple forces such as forcings arising from pressure gradients, Coriolis force, gravity and buoyancy. In nature, we observe that atmospheric flow is shaped by a dominant equilibrium among a subset of these forces, where the magnitudes of these dominant terms are significantly larger than those of other terms in the equations. To understand atmospheric balance principles, one must analyse the system of differential equations that govern the atmospheric evolution. Nondimensionalisation and scaling help to identify predominant balances and the conditions under which they apply.

This subsection provides an explanation of a set of atmospheric equations, their nondimensionalisation, and three key balance principles: hydrostatic, geostrophic, and weak temperature gradient balances. Each balance principle is associated with a derived scale number that characterises the circumstances in which it holds. These scale numbers facilitate the identification of different dynamical regimes based on the considered space and time scales. After explaining these dynamical regimes, we will discuss deviations from the predominant balances, clarifying imbalanced motion. The theory presented in this section primarily draws from the studies conducted by Craig and Selz (2018) and Klein (2010). The foundational principles of nondimensionalisation and scaling adhere to the formulations provided by Vallis (2017).

Equations of Motion As in Craig and Selz (2018), we use the anelastic equations for our explanations, as these are a simplification of the full Navier-Stokes equations, but still include all the dynamic processes relevant to this work. More precisely, the anelastic approximation removes acoustic waves, which do not interact with the atmospheric flow.

The anelastic equations consider perturbations from a horizontally uniform reference state in hydrostatic balance. Hence, the system state is described through the following variables:

$$\varrho(x,y,z,t) = \varrho_0(z) + \varrho_1(x,y,z,t), \tag{2.31}$$

$$\pi(x,y,z,t) = \pi_0(z) + \pi_1(x,y,z,t), \tag{2.32}$$

$$\theta(x,y,z,t) = \theta_0(z) + \theta_1(x,y,z,t), \tag{2.33}$$

where ϱ denotes the density, π denotes the Exner pressure, and θ the potential temperature. (x,y,z) is the location and t denotes the time. In the above equations, variables with index 0 indicate the horizontally uniform reference state and variables with index 1 indicate the perturbations. It is assumed that the size of the perturbation is small with respect to the reference state, i.e. $|a_1| \ll |a_0|$ with $a \in \{\varrho, \pi, \theta\}$. The definitions of Exner pressure and potential temperature are given by

$$\pi = \left(\frac{p}{p_0}\right)^{\frac{R}{c_p}}, \qquad \theta = \left(\frac{p_0}{p}\right)^{\frac{R}{c_p}}, \tag{2.34}$$

with p being pressure, R being the gas constant for dry air and c_p being the specific heat capacity at constant pressure. For convenience, to make the equations more readable, a rescaled pressure variable is introduced, $\Phi = c_p\theta_0\pi_1$. With this the anelastic equations are:

$$\underbrace{\partial_t \mathbf{v} + \mathbf{u}\cdot\nabla\mathbf{v}}_{\text{advective derivative}} + \underbrace{f\hat{k}\times\mathbf{v}}_{\text{coriolis force}} + \underbrace{\nabla_h\Phi}_{\text{hor. pressure gradient}} = 0, \tag{2.35}$$

$$\underbrace{\partial_t w + \mathbf{u}\cdot\nabla w}_{\text{advective derivative}} + \underbrace{\partial_z\Phi}_{\text{vert. pressure gradient}} = \underbrace{\frac{g}{\theta_{00}}\theta_1}_{\text{buoyancy}}, \tag{2.36}$$

$$\underbrace{\nabla_h\cdot\mathbf{v}}_{\text{hor. wind divergence}} + \underbrace{\rho_o^{-1}\partial_z(\rho_0 w)}_{\text{change in vert. wind}} = 0, \tag{2.37}$$

$$\underbrace{\partial_t\theta_1 + \mathbf{u}\cdot\nabla\theta_1}_{\text{advective derivative}} + \underbrace{w\partial_z\theta_0}_{\text{vert. advection of buoyancy}} = \underbrace{q_\theta}_{\text{heating}} \tag{2.38}$$

with ∂_t and $\nabla = (\partial_x \quad \partial_y \quad \partial_z)^T$ being partial derivatives in time and space and $\nabla_h = (\partial_x \quad \partial_y)^T$ indicating the horizontal gradient. $\mathbf{u} = (u \quad v \quad w)^T$ denotes the three-dimensional wind vector and $\mathbf{v} = (u \quad v)^T$ the horizontal wind vector. θ_{00} is a constant reference surface potential temperature. f is the Coriolis parameter. q_θ is the diabatic potential heating rate per unit mass. Eq. (2.35) and Eq. (2.36) are the horizontal and vertical momentum equations, respectively. Eq. (2.37) is the continuity equation and Eq. (2.38) is the thermodynamic equation. We will refer to the individual terms in the equations as indicated below the underbraces. Note that these equations allow for internal gravity waves through the interplay of vertical advection of the background gradient of potential temperature in Eq. (2.38), which is related to a change in buoyancy, and the buoyancy term on the right hand side of Eq. (2.36).

Nondimensionalisation and Scaling It is possible to identify predominant balances in the equation of motions by comparing relative magnitudes of the individual terms of the equations. Naturally, the magnitudes of the terms depend on the typical spatial and temporal scales of the considered problem. Choosing sensible scaling parameters is therefore essential. By nondimensionalising, the variables in the anelastic equations can be represented as multiples of their typical scales. In mathematical terms this looks like

$$t = \hat{t}T \qquad \mathbf{x} = \hat{\mathbf{x}}L \qquad z = \hat{z}H, \tag{2.39}$$

$$\mathbf{v} = \hat{\mathbf{v}}U \qquad w = \hat{w}W \qquad \theta = \hat{\theta}\Theta, \tag{2.40}$$

where variables with a hat are the nondimensional variables and here $\mathbf{x} = (x \quad y)^T$. The capital letters denote the typical scales which depend on the atmospheric phenomenon that is considered. In the following, we will use nondimensionalisation and scaling to explain three fundamental balance principles particularly relevant to this work.

Hydrostatic Balance considers the aspect ratio of the flow. Nondimensionalisation of the vertical momentum equation (Eq. 2.36) yields

$$\frac{UW}{L}(\partial_{\hat{t}}\hat{w} \;+\; \hat{\mathbf{u}} \cdot \nabla\hat{w}) \;+\; \frac{U^2}{H}\partial_{\hat{z}}\hat{\Phi} \;=\; \left|\frac{g}{\theta_{00}}\theta_1\right|, \tag{2.41}$$

here we assume that all derivatives (∇, ∂) that act on nondimensional variable are also nondimensional. As a typical timescale T we use the advective time scale with $T_{adv} = \frac{L}{U}$. Moreover, nondimensionalisation of the horizontal momentum equation (Eq. 2.35) yields $\Phi \sim U^2$. Nondimensionalisation of the continuity equation (Eq. 2.37) yields $\frac{W}{H} \sim \frac{U}{L}$. Substituting this into Eq. (2.41) results in

$$\frac{U^2H}{L}(\partial_{\hat{t}}\hat{w} \;+\; \hat{\mathbf{u}} \cdot \nabla\hat{w}) \;+\; \frac{U^2}{H}\partial_{\hat{z}}\hat{\Phi} \;=\; \left|\frac{g}{\theta_{00}}\theta_1\right|. \tag{2.42}$$

Comparing the advective with the pressure gradient term yields,

$$\frac{\|\partial_t w \;+\; \mathbf{u} \cdot \nabla w\|}{\|\partial_z \Phi\|} \sim \left(\frac{H}{L}\right)^2 = \alpha^2 \tag{2.43}$$

where α denotes the aspect ratio. If the vertical length scale is much smaller compared to the horizontal length scale the aspect ratio is small and hence the advective term can be neglected and the pressure gradient term balances the gravitational force:

$$\partial_z \Phi \;=\; \frac{g}{\theta_{00}}\theta_1 \tag{2.44}$$

This equation is called hydrostatic balance. In other words, hydrostatic balance means that the pressure at a given point is determined solely by the weight of the air column above the this point, and the vertical velocities within this column are negligible.

Geostrophic Balance can be derived by scale analysis of the horizontal momentum equation (Eq. 2.35), that is

$$\frac{U^2}{L}(\partial_{\hat{t}}\hat{\mathbf{v}} \; + \; \hat{\mathbf{u}}\cdot\nabla\hat{\mathbf{v}}) \; + \; fU\hat{k}\times\hat{\mathbf{v}} \; + \; \nabla_h\Phi \; = \; 0 \tag{2.45}$$

Comparing the advective with the rotational term yields

$$\frac{\|\partial_t\mathbf{v} \; + \; \mathbf{u}\cdot\nabla\mathbf{v}\|}{\|f\hat{k}\times\mathbf{v}\|} \sim \frac{U}{fL} = R_o, \tag{2.46}$$

where R_o denotes the Rossby number. If the ratio between the horizontal wind velocity and the rotational speed is very small, the Rossby number is very small. In this case the advective term can be neglected and the Coriolis force balances the horizontal pressure gradient force:

$$fk\times\mathbf{v} \; = - \; \nabla_h\Phi. \tag{2.47}$$

This equation is called geostrophic balance.

Weak Temperature Gradient Balance can be derived by scale analysis of the thermodynamic equation (Eq. 2.38). However, it is more convenient to nondimensionalise the equation, introducing $N^2 \; = \; \frac{g}{\theta_0 0}\partial_z\theta_0$, where N^2 is called the buoyancy or Brunt-Väisäla frequency. N^2 reflects the characteristic frequency of short internal waves. With that Eq. (2.37) can be nondimensionalised as

$$\frac{U\Theta_1}{L}(\partial_{\hat{t}}\hat{\theta}_1 + \hat{\mathbf{u}}\cdot\nabla\hat{\theta}_1) \; + \; \frac{\theta_{00}}{g}WN^2(\hat{w}\hat{N}^2) \; = \; q_\theta. \tag{2.48}$$

with $N^2 = \bar{N}^2\hat{N}^2$. From Eq. (2.41) follows $\Theta_1 \sim \frac{\theta_{00}}{g}\frac{U^2}{H}$. And, similar as in the scaling for hydrostatic balance, we use Eq. (2.37) to estimate $W \sim \frac{UH}{L}$.
Comparing the advective term with the vertical advection of the background gradient yields

$$\frac{\|\partial_t\theta_1 \; + \; \mathbf{u}\cdot\nabla\theta_1\|}{\|w\partial_z\theta_0\|} = \frac{U^3}{LHN^2W} = \frac{U^2}{N^2H^2} = Fr^2, \tag{2.49}$$

where Fr denotes the Froude number. If the advective forcing is much smaller than the buoyancy forcing, the Froude number is very small and the advective term can be neglected, hence

$$w\partial_z\theta_0 = q_\theta. \tag{2.50}$$

This equation is called Weak Temperature Gradient (WTG) balance. Note that, since the advection of θ_1 is neglected, the WTG balance suppresses gravity waves, as the term on the right hand side in Eq. (2.36) is constant in time.

Dynamical Regimes Using typical scale values for atmospheric phenomena, one can infer, through the scaling parameters, the predominant force balance that determines a given phenomenon. If, for instance, we consider a low-pressure system, it typically spans a spatial scale of $L = 1000$ km, with a typical horizontal wind speed of $U = 10\ \frac{\mathrm{m}}{\mathrm{s}}$, and the Coriolis parameter in mid-latitudes is on the order of $f = 10^{-4}\ \frac{1}{\mathrm{s}}$. This yields a Rossby number of $R_0 = 0.1$, indicating that the cyclone is in geostrophic balance, which means that the pressure gradient force is in equilibrium with the Coriolis force.
On the mesoscale, we can assume $L = 100$ km, $W = 1\ \frac{\mathrm{m}}{\mathrm{s}}$, $U = 10\ \frac{\mathrm{m}}{\mathrm{s}}$, $H = 10$ km (the approximate scale height of the atmosphere) and $N = 10^{-2}\frac{1}{\mathrm{s}}$, with f remaining unchanged, naturally. This results in $R_0 = 1$ and $Fr^2 = 1$. Consequently, on the mesoscale, Geostrophy seems to be as good of an approximation as the weak temperature gradient balance and it does not seem straight forward which balance principle should be used. Models that employ the weak temperature gradient approximation are frequently used to model near-equatorial flows, since deep-convective processes associated with strong diabatic heating rates that build up play an important role in the tropics, yielding $Fr^2 < 1$ (see e.g. Klein, 2006; Majda and Klein, 2003).
In their study, Craig and Selz (2018) divided the mesoscale itself into different dynamic regimes based on the calculation of scale numbers from convective-resolving numerical simulations in the midlatitudes. Figure 2.3 shows the computed space-time spectra from their study. The color in each panel indicates the size of the dimensionless parameters. The straight dashed line in all three panels corresponds to the wave speed of 10 $\frac{m}{s}$ which is a classical wave speed for advective motions. The convective scale, encompassing the smallest scale within the mesoscale, is particularly relevant for this study. With characteristic length and time scales on the order of $L = 10$ km and $T = 1$ hour, respectively, it closely approaches the unresolved scales as indicated by the hatched

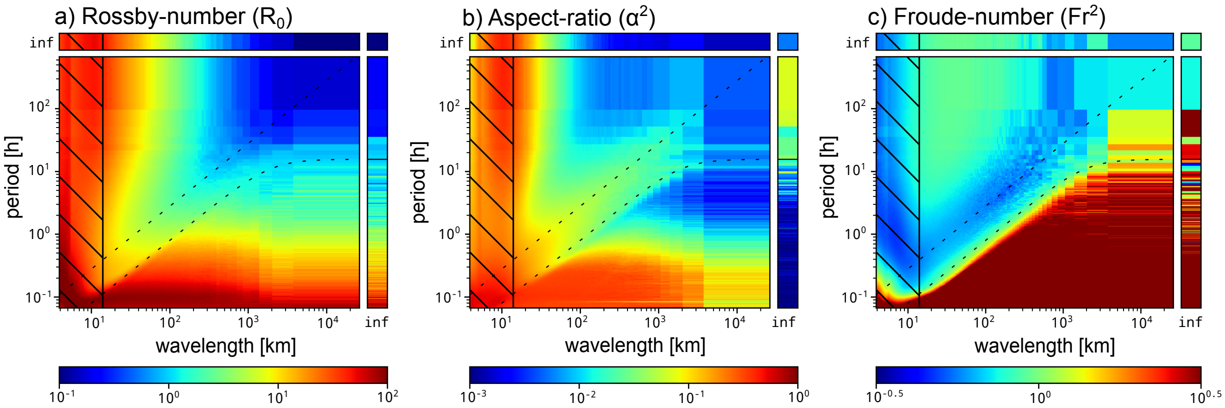


Figure 2.3: Space-time spectrum of a convective-resolving mid-latitudinal weather simulation showing the nondimensional parameters given in Eq. 2.43, Eq. 2.46 and Eq. 2.49 in 5 km height. The colored bands above and to the right of the main plots represent the modes where the wavelength is infinity (right), the time period is infinity (top), or both (top right corner). The hatched areas indicate spatial modes where the wavelength is close to the grid length of the simulation. The upper dotted (diagonal) line refers to a typical advection speed of 10 $\frac{\text{m}}{\text{s}}$ while the lower line gives the dispersion relation for deep, horizontally propagating gravity waves. For more details see Craig and Selz (2018, Figure 3). This figure and parts of the caption are adopted from there.

areas in Figure 2.3. However, it is still notable in Panels a and b that for these scales, $R_0 \approx 10$ and $\alpha^2 \approx 1$. This means that hydrostatic and geostrophic balance do not represent good approximations, which is in accordance with our derived scale parameters. Panel c indicates $Fr^2 \approx 0.5$ for the convective scale. This means that the weak temperature gradient approximation might be applicable, even though one should note that it is not as good an approximation as geostrophic balance on the synoptic scale. For more details the reader is referred to Craig and Selz (2018).

Imbalanced Motion While predominant balances and dynamical regimes have profound implications for the atmospheric flow, they are always an approximation, and weak imbalanced motion will always be present. Moreover, as already mentioned, DA can excite unrealistic imbalanced motions. Nevertheless, an imbalanced flow will always strive to evolve into a balanced flow. Understanding the characteristics of these imbalanced motions and the mechanisms by which the atmosphere relaxes into a balanced state is essential for developing methods to diagnose or mitigate undesired imbalance. This paragraph aims to elucidate this process, using the example of geostrophic adjustment, again following Vallis (2017).

The easiest way to explain geostrophic adjustment is to use the Shallow Water Equations (SWE). SWE represent a simplified version of the Navier-Stokes equations, describing the evolution of a fluid layer with constant density and a small aspect ratio. The movement of the fluid is entirely captured by momentum and mass continuity equations. The variables simulated include the displacement of the free fluid surface and horizontal flow velocities. Despite its simplicity, this model allows for an explanation of the fundamental mechanisms of geostrophic adjustment without loss of generality of the concept.

Let's assume that the initial state of the fluid is at rest but exhibits a discontinuity in the height field in the form of a step function. Such an initial condition could be created by separating two water masses of different depths with a thin barrier and then quickly removing that barrier. In this process, the initial displacement of the water surface should remain small compared to the horizontal scale. In Figure 2.4, this initial state is represented by the red dashed line, with the step function in the height field and zero horizontal velocities (also shown in Figure 2.5, Panel a). The blue line in Figure 2.4 indicates the evolution of the fluid layer at a later point in time, shortly after removing the thin barrier. The left column of Figure 2.4 shows the evolution of the fluid layer in a rotating system, i.e. with the Coriolis force term in the momentum equation and the right column shows the evolution of the fluid layer in a non-rotating system, so no Coriolis force in the momentum equation. In both cases, one can observe that the initial pattern in the height field moves to the left and right in the form of a gravity wave front. In the non-rotating case, after the front has passed, the fluid height is again undisturbed (at zero in Fig. 2.4) and a uniform, non-zero velocity is established in the u-direction. In the rotating case the initial disturbance is not radiated outward to infinity completely as in the non-rotating problem. Rather, the adjustment of the initially sharp gradient in the height field is constrained up to a certain distance, which is often referred to the Rossby Radius of Deformation (L_d), leaving behind a smoother gradient. The reason for this is that in the final (equilibrium) state the Coriolis force balances the pressure gradient force caused by the non-zero gradient in the fluid height. The initial and final height fields in the rotating case are depicted in Panels a and b of Figure 2.5. The geostrophically balanced flow velocity is perpendicular to the slope of the free surface as depicted in Figure 2.5, Panel c. In both cases, rotating and non-rotating, the adjustment from an initial unbalanced state, characterised by a sharp gradient in the height field, to a balanced state is achieved by transient gravity waves.

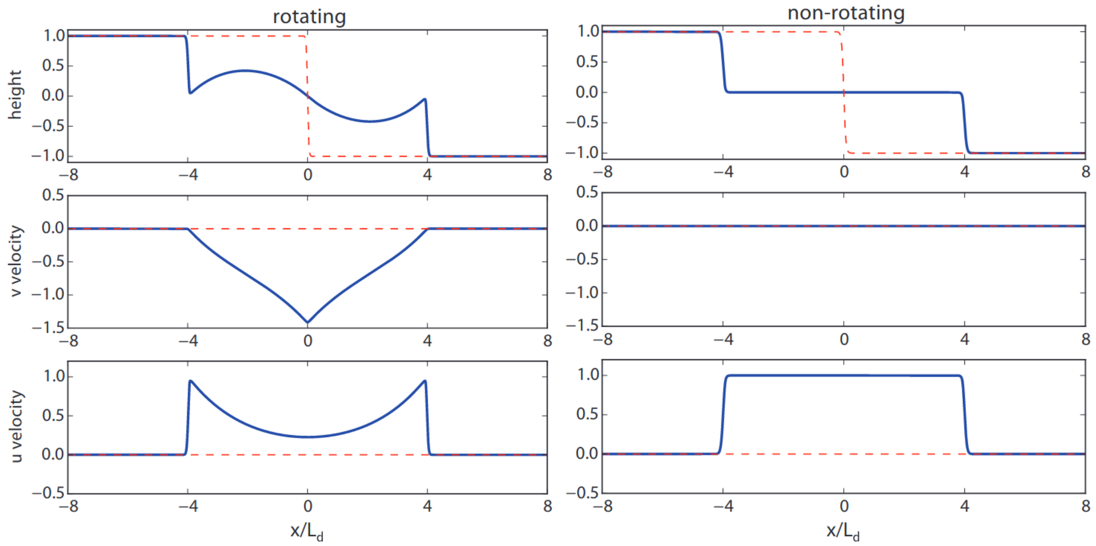


Figure 2.4: The solution of the shallow water equations with and without rotation obtained by numerical integration. The panels show snapshots of the state of the fluid (solid blue lines) after being released from the stationary initial state (red dashed lines). In the rotating case the flow will evolve toward an end state in geostrophic balance, whereas in the non-rotating case the flow will eventually become stationary. In the non-rotating case L_d is defined with the rotating parameters. Figure and caption are adopted from Vallis (2017, Figure 3.11).

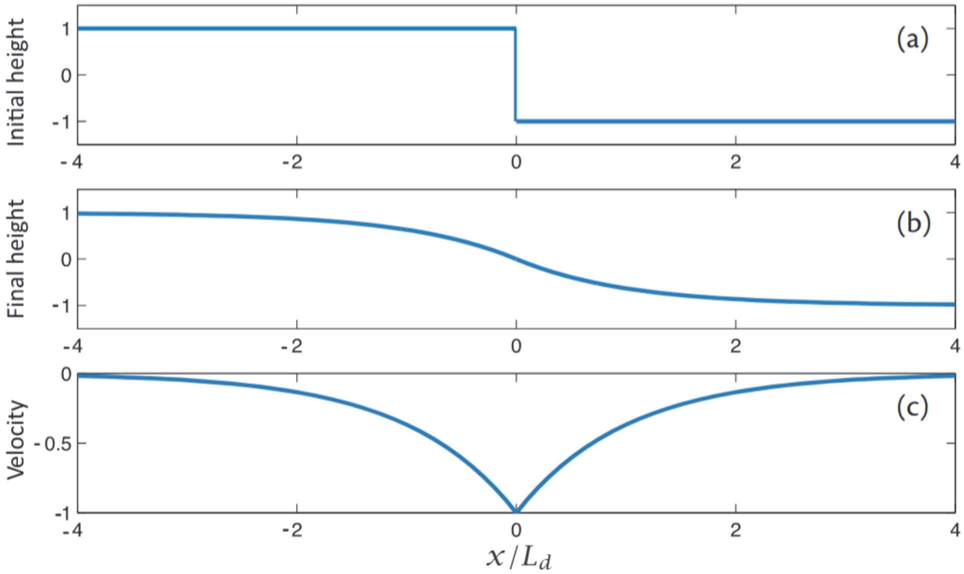


Figure 2.5: The initial and end state of a linear geostrophic adjustment problem (rotating case). a) Initial height field, b) Equilibrium (final) height field, c) Equilibrium geostrophic velocity, normal to the gradient of height field. Figure and caption are adopted from Vallis (2017, Figure 3.10).

2.2.2 Imbalance in Data Assimilation

Data assimilation has the potential to disrupt prevailing balances established by model dynamics. The imbalance introduced through data assimilation can be thought of as a discontinuity in one or more model variables, and similarly to what was discussed in the previous section, it can be expected that the model will attempt to re-establish equilibrium through adjustment processes in the form of transient gravity waves. The following discussion will address why the data assimilation may produce sharp gradients in the analysis and explore potential practical consequences of these imbalances. We will start with the reasons behind imbalance.

As discussed, the analysis produced by the LETKF is a linear combination of the background ensemble members. Each background ensemble member is a solution of the system of nonlinear differential equations that describes the temporal evolution of the atmosphere. Since the system is nonlinear a linear combination of the ensemble of solutions is not necessarily a new solution to the system. Thus, the assimilation of observations may yield analyses that are dynamically inconsistent with the forecasting model. Nevertheless, it is important to note that the model error covariance matrix $\mathbf{P}^b$ should capture physical correlations and balances and that linear balances could be preserved in linear combinations of the system. Therefore, the imbalance introduced solely through the linear combination might not be the biggest source of imbalance.

The major source of imbalance is localisation. Localisation is an ad-hoc method that is needed to solve high-dimensional problems such as estimating the current state of the atmosphere. Localisation essentially means that for each grid-point in the analysis field a different set of weights is computed and thus a different linear combination of first-guess ensemble members is assembled (Greybush et al., 2011). This implies that unphysical gradients may arise. Another interesting aspect that follows is that a small analysis error, i.e. an analysis that is very close to the observations does not guarantee a dynamically balanced state Lange and Craig (2014).

Studies (e.g. Aksoy et al., 2010; Greybush et al., 2011; Lange et al., 2017) suggest that abundance of gravity motion, introduced through DA, may trigger spurious convective cells that have the potential to degrade the forecast. The gravity wave displaces air parcels from their equilibrium position. The stability of the atmospheric stratification then determines whether the air parcel continues to rise, forming a cloud or even a convective cell, or if the air parcel swings back to its initial position. Figure 2.6 shows an example of spurious convection. The figure is extracted from Lange and Craig (2014), where they conducted idealised data assimilation experiments using the LETKF with a relatively small localisation radius. It shows snapshots of radar reflectivity. In Panel a) the model simulation without data assimilation, commonly referred to as the nature run, is depicted. This simulation involves a supercell characterised by high precipitation, evident in high values of radar reflectivity. The subsequent panels in Figure 2.6 illustrate the analysis ensemble mean and various analysis ensemble members. It is noticeable that the supercell itself is well-captured in the analysis; however, scattered features around the cell represent spurious cells associated with imbalance and spurious gravity waves. Lange and Craig (2014) as well as others have shown that the imbalance seen in the analysis and the spin-up of the forecast is sensitive to assimilation settings, such as the cycling length, the localisation length scale and the horizontal averaging of observations. Given the potential detrimental effect of imbalance introduced through data assimilation on the forecast it becomes evident that the goal of data assimilation itself should be to produce an initial state that is well-balanced. Therefore, it is important to understand and diagnose imbalance. In Chapter 4 we present how we have implemented and tested different forms of imbalance metrics that might be relevant to a convective-scale DA system.

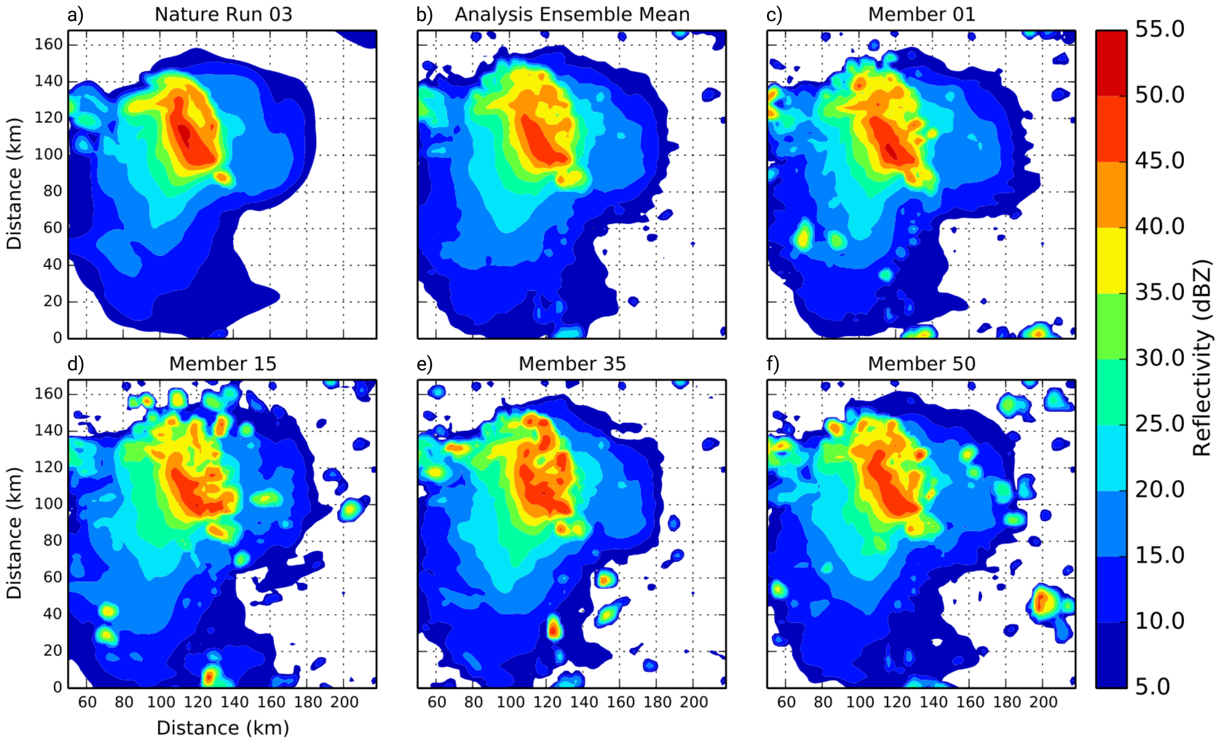


Figure 2.6: Composite reflectivity of a nature run (Panel a), the analysis ensemble mean (Panel b) and other individual ensemble members (Panels c-f) at analysis time. The figure is taken from Lange and Craig (2014, Figure 11). The snapshots depict a super cell and spurious convection or noise around the super cell in case of data assimilation.

Chapter 3

Partial Analysis Increments as Diagnostic for LETKF Data Assimilation Systems

Convective-scale data assimilation requires frequent analysis updates, using a multitude of spatially and temporally highly resolved observations. While observational data sources, such as geostationary satellites, or ground based remote sensing instruments are promising sources of information in this regard, their assimilation is complex and requires a number of approximations that do not always reflect reality. Consequently, it is important to assess the analysis influence of observations to make sure their assimilation is beneficial. In this chapter, a new diagnostic for quantifying analysis influence of observations in a Local Ensemble Transform Kalman Filter is presented, thereby addressing Research Question 1: *How can we evaluate the impact of various observations in a way that yields easily interpretable results and ensures computational efficiency?*

This chapter is based on: Diefenbach, T., Craig, G., Keil, C., Scheck, L. & Weissmann, M.(2023) Partial analysis increments as diagnostic for LETKF data assimilation systems. Quarterly Journal of the Royal Meteorological Society, 149(752), 740–756. Available from: `https://doi.org/10.1002/qj.4419`

3.1 Recap and State of the Art

Ensemble data assimilation systems such as the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) have become a well-established approach for regional, convection-permitting numerical weather prediction (NWP) models as they are both computationally efficient and include flow-dependent estimates of error covariances. This is, for example, reflected in the operational implementation of an LETKF data assimilation system in the regional NWP system of Deutscher Wetterdienst (DWD) in 2017 (Schraff et al., 2016). Due to computational restrictions, the ensemble size is, however, usually restricted, which introduces spurious correlations and the need for covariance localisation (Necker et al., 2020a,b). Furthermore, the LETKF minimizes the cost function locally in observation space, which introduces difficulties for the assimilation of non-local satellite observations that provide vertically integrated information on atmospheric constituents emitting or scattering radiation. Non-linearity, non-Gaussianity, systematic model deficiencies in the representation of hydrometeors and their radiative properties (Geiss et al., 2021) as well as significant uncertainty of radiative transfer models in cloudy situations (Scheck et al., 2018) add further complexity to the assimilation of cloud-affected satellite observations in convection-permitting assimilation systems (Hu et al., 2022). Nevertheless, these observations provide potentially very valuable information for convective-scale data assimilation (Gustafsson et al., 2018; Schroettle et al., 2020) and their assimilation is therefore a very active area of research (Okamoto, 2017; Scheck et al., 2020; Zhang et al., 2016). To overcome these difficulties for the assimilation of cloud-affected satellite radiance observations, several studies (e.g. Schomburg et al., 2015; Scheck et al., 2020; Bauer et al., 2010) conducted single-observations experiments to better understand the influence of such observations in data assimilation systems. Such experiments, however, require to run a full, computationally expensive data assimilation experiment for the assimilation of just one observation or a very limited number of spatially well-separated observations that do not influence each other in the assimilation process. In this chapter, we propose a significantly more efficient approach for investigating the three-dimensional analysis influence of individual observations (partial analysis increment, PAI) based on available LETKF analysis ensemble perturbations. This new diagnostic for partial analysis increments related to a single observation allows to approximate the contribution of individual observations to the analysis or the contribution that an observation would have with modified assimilation settings (e.g. modified assigned localisation scale or observation error).

The strength of the diagnostic is that it allows for investigating the 3D structure of the analysis contribution of one observation directly in model space. By that it is possible to detect where observations draw the analysis into opposite directions, which is especially interesting with respect to the assimilation of novel observations. While detrimental observation influence is part of the statistical nature of the data assimilation (especially when the model state is already very close to the truth) patterns or large values of detrimental observation influence in the analysis may be an indication for suboptimal data assimilation related to e.g. spurious correlations or wrong localization settings. Other existing diagnostics such as observation influence (Liu et al., 2009; Cardinali et al., 2004) focus on the relative contribution of observations to the analysis as dimensionless scalar quantities. Furthermore, several studies used ensemble forecast sensitivity to observations (EFSO, Kalnay et al., 2012; Sommer and Weissmann, 2014, 2016; Kotsuki et al., 2019), to approximate forecast observation impact of individual observations in a computationally cheap way, without running multiple experiments. While in principle partial analysis increments are included in the derivation of EFSO (c.f. Ota et al., 2013; Hotta et al., 2017a) the focus in these EFSO-studies is mainly on the statistical contribution of observations to the reduction of forecast error with usual lead times of $\mathcal{O}(\text{hours})$. The PAI diagnostic is limited to the investigation of analysis influence, but it has the advantage that it avoids inaccuracies related to the linearity assumption of the forecast evolution, issues with localisation of the forecast error and the verification of the forecast error (Necker et al., 2018) in contrast to EFSO. The objective of this study is to show that the PAI diagnostic can be used as an economical alternative to single observation experiments and as a diagnostic to evaluate and even optimise the data assimilation system. Moreover, the derivation of PAI is given in detail with a special focus on the approximations that have to be made to apply it to a near-operational LETKF data assimilation system.

The remainder of this chapter is structured as follows: Section 3.2 presents the detailed derivation of the PAI methodology as well as a description of the modelling and assimilation system, the experimental setup and the applied metrics. In Section 3.3, we illustrate PAI results for several examples and discuss the effect of the approximations by comparison of PAI results with the analysis influence in single-observation experiments. Section 3.4 presents three potential applications of the PAI diagnostic, namely the analysis of the contribution of different observations to analysis fields, the detection of detrimental observation influence and the optimization of assimilation settings. Finally, conclusions are provided in Section 3.5.

3.2 Method and Data

In this study, we employ the Km-scale ENsemble Data Assimilation (KENDA) system of DWD (Schraff et al., 2016). The KENDA system comprises an LETKF assimilation scheme (after Hunt et al. (2007)) that is coupled with a non-hydrostatic regional NWP model (in this study the COnsortium for Small-scale MOdeling (COSMO) model). The LETKF provides the analysis ensemble in a computationally efficient way by transforming the problem from high-dimensional model space into low-dimensional ensemble space and by computing the analysis locally on a reduced analysis grid. The localisation not only makes the method more efficient, but is also necessary to mitigate spurious correlations and increase the degrees of freedom of the analysis. In the following we will derive the mathematical formulation for PAI from the LETKF equations and describe the approximations that are involved. In the derivation we will use the same notation as in Hunt et al. (2007).

3.2.1 Partial Analysis Increment Formulation

Before getting to the the PAI formulation for LETKF systems, we start with the general form of the sequential analysis equation, where the analysis $\boldsymbol{x}^a$ is produced by a statistical combination of the background $\boldsymbol{x}^b$ and the observations $\boldsymbol{y}^o$ (see eg. Eq. (5.4.1) in Kalnay, 2003):

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \mathbf{K}\left[\boldsymbol{y}^o - H(\boldsymbol{x}^b)\right]. \tag{3.1}$$

H denotes the non-linear observation operator, which transforms a vector from n-dimensional model space into p-dimensional observation space. The term $\mathbf{K}$ is often referred to as the Kalman gain matrix. The analysis increment is defined as the difference between the analysis and the background:

$$\delta\boldsymbol{x} = \boldsymbol{x}^a - \boldsymbol{x}^b = \mathbf{K}(\boldsymbol{y}^o - \boldsymbol{y}^b). \tag{3.2}$$

$(\boldsymbol{y}^o - \boldsymbol{y}^b)$ is called the innovation vector or background departure, with $\boldsymbol{y}^b = H(\boldsymbol{x}^b)$ being the model equivalent of the observations. From this expression it becomes clear that

$\mathbf{K}$ is a matrix of dimension n x p that determines the weight of the correction and transforms back from observation space to model space. Assuming that $\mathbf{K}$ is known, the formulation of partial analysis increment (PAI) is straight-forward from Eq. (3.2). The partial analysis increment which is related to one single observation, $\boldsymbol{y}_j^o$, is then defined as:

$$\delta \boldsymbol{x}_j = \mathbf{K}_j \, (\boldsymbol{y}^o - \boldsymbol{y}^b)_j, \tag{3.3}$$

where the index j is used to indicate that only the j-th column of $\mathbf{K}$ and the j-th row of the innovation vector are considered. The sum over all PAIs equals the total increment, i.e.

$$\sum_{j=1}^{p} \delta \boldsymbol{x}_j = \delta \boldsymbol{x}. \tag{3.4}$$

Similarly, it is possible to calculate PAIs for subsets of observations which is simply the sum of all partial increments of all observations in the subset.

However, in practice this formulation cannot be used directly since in the LETKF the analysis is carried out in ensemble space and $\mathbf{K}$ is never calculated explicitly. It is possible, though, to express $\mathbf{K}$ in terms of standard LETKF output data products as:

$$\mathbf{K} = (k-1)^{-1} \, \mathbf{X}^a \mathbf{Y}^{aT} \mathbf{R}^{-1}, \tag{3.5}$$

where $\mathbf{X}^a$ and $\mathbf{Y}^a$ are the ensemble analysis perturbation matrices in model space and observation space, respectively and k is the number of ensemble members. This formulation of $\mathbf{K}$ has been used before in the context of observation influence by other studies such as for example Kalnay et al. (2012) or Hotta et al. (2017a) and can also be found in Eq. (9) in Gustafsson et al. (2018). In the study of Kalnay et al. (2012), the derivation of Eq. (3.5) assumes a linear observation operator $\mathbf{H}$. In the following we will derive Eq. (3.5) from the LETKF equations for a non-linear H, using the linear approximation in ensemble space (Eq. 18 in Hunt et al., 2007) that is also employed for the computation of the analysis.

The LETKF approximates the background and analysis uncertainty by an ensemble and computes the analysis ensemble mean $\overline{\boldsymbol{x}}^a$ as an optimal linear combination of the background ensemble members. The analysis equation for the LETKF is (cf. Eq. (22) in Hunt et al. 2007):

$$\overline{\boldsymbol{x}}^a = \overline{\boldsymbol{x}}^b + \mathbf{X}^b \overline{\boldsymbol{w}}^a. \tag{3.6}$$

where $\overline{\boldsymbol{w}}^a$ is the weight vector that minimises the LETKF-cost function in ensemble space. The overbars indicate the ensemble mean. The $n \times k$ matrix $\mathbf{X}^b$ is the background ensemble perturbation matrix. Column i of $\mathbf{X}^b$ is defined as $\boldsymbol{x}^b_{(i)} - \overline{\boldsymbol{x}}^b$, i.e. the deviation of one ensemble member i from the ensemble mean. From Eq. (20) of Hunt et al. (2007) we know that

$$\overline{\boldsymbol{w}}^a = \tilde{\mathbf{P}}^a \mathbf{Y}^{bT} \mathbf{R}^{-1} (\boldsymbol{y}^o - \boldsymbol{y}^b), \tag{3.7}$$

where $\tilde{\mathbf{P}}^a$ is the analysis error covariance matrix in ensemble space, $\mathbf{Y}^b$ is the background ensemble perturbation matrix transformed into observation space (with dimensions p x k) and $\mathbf{R}$ is the observation error covariance matrix.

The individual ensemble members $\boldsymbol{x}^a_{(i)}$ are distributed around the ensemble mean such that their spread reflects the uncertainty of the analysis in ensemble space ($\tilde{\mathbf{P}}^a$), which can be computed explicitly. The weight vectors for the individual ensemble members $\boldsymbol{w}^a_{(i)}$ are chosen as the symmetric square root of $\tilde{\mathbf{P}}^a$:

$$\tilde{\mathbf{P}}^a = (k-1)^{-1}\, \mathbf{W}^a \mathbf{W}^{aT}, \tag{3.8}$$

with $\mathbf{W}^a$ being the ensemble weight perturbation matrix in ensemble space, with columns $\boldsymbol{w}^a_{(i)} - \overline{\boldsymbol{w}^a}$. Hence, the individual ensemble members are given as:

$$\boldsymbol{x}^a_{(i)} = \overline{\boldsymbol{x}}^b + \mathbf{X}^b (\mathbf{W}^a_{(i)} + \overline{\boldsymbol{w}}^a) \tag{3.9}$$

Taking the difference between Eq. (3.6) and Eq. (3.9) shows that the analysis ensemble perturbations are given as:

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}^a. \tag{3.10}$$

If we now insert Eq. (3.7), Eq. (3.8) and Eq. (3.10) into Eq. (3.6) we get:

$$\overline{\boldsymbol{x}}^a = \overline{\boldsymbol{x}}^b + (k-1)^{-1}\, \mathbf{X}^a \mathbf{W}^{aT} \mathbf{Y}^{bT} \mathbf{R}^{-1} (\boldsymbol{y}^o - \boldsymbol{y}^b) \tag{3.11}$$

Instead of linearising H around the ensemble mean, which would involve a large $p \times n$ Jacobian matrix, Hunt et al. (2007) make a linear approximation in ensemble space to relate perturbations in model state to observations space,

$$H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b \boldsymbol{w}) \approx \overline{\boldsymbol{y}}^b + \mathbf{Y}^b \boldsymbol{w} \tag{3.12}$$

(see Eq. 18 in Hunt et al., 2007). Using the same assumption,

$$\boldsymbol{y}^a(i) - \overline{\boldsymbol{y}}^a = H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b \boldsymbol{w}^a_{(i)}) - H(\overline{\boldsymbol{x}}^b + \mathbf{X}^b \overline{\boldsymbol{w}}^a) = \mathbf{Y}^b(\boldsymbol{w}^a_{(i)} - \overline{\boldsymbol{w}}^a), \tag{3.13}$$

and hence $\mathbf{Y}^a = \mathbf{Y}^b \mathbf{W}^a$. Inserting this into Eq. (3.11) yields the desired expression for $\mathbf{K}$ as given in Eq. (3.5):

$$\overline{\boldsymbol{x}}^a = \overline{\boldsymbol{x}}^b + (k-1)^{-1}\, \mathbf{X}^a \mathbf{Y}^{aT} \mathbf{R}^{-1} (\boldsymbol{y}^o - \boldsymbol{y}^b). \tag{3.14}$$

In this context, it should be noted that the linear approximation (Eq. 3.12) leads to a suboptimal analysis in case of non-linear observation operators. The PAI diagnostic described above, however, is consistent with the assumption of the LETKF and therefore reflects the actual analysis increment (apart from the approximation related to localisation discussed below).

So far we ignored the effects of localisation although it is a crucial part of the LETKF. Localisation means that the analysis is carried out independently for the individual model grid points (or on a reduced grid as in KENDA). This is achieved by considering only the observations in a certain region around the location of the respective grid point for the analysis weight calculation. To achieve a smooth and physically consistent analysis, neighbouring analysis points should largely use the same set of observations and the influence of distant observations is reduced gradually. In a mathematical sense, this means that the elements of $\mathbf{R}^{-1}$ are multiplied by a weighting factor, which is equal to one at the location of the analysis and decays to zero after a certain radius. Reducing elements of $\mathbf{R}^{-1}$ means increasing the assumed observation error and thus giving less weight to the respective observation. The weighting function used in the LETKF is the Gaspari-Cohn function, which is a Gaussian-shaped curve that decays to zero after the so-called cut-off radius. The cut-off radius is defined as $r = 2\ell\sqrt{\frac{10}{3}}$, where ℓ is called the localisation length scale. For the PAI diagnostic, localisation has two implications:

1. The analysis in the LETKF is computed using a localised $\mathbf{R}$. Thus, for a diagonal $\mathbf{R}$, which is used in the KENDA system and throughout this study, the localized version of the Kalman gain from Eq. (3.5) can be written as follows:

$$\mathbf{K}_{loc} = (k-1)^{-1} \mathbf{X}^a \mathbf{Y}^{aT} \mathbf{R}^{-1} \circ \rho, \tag{3.15}$$

where ρ is a matrix of Gaspari-Cohn factors and $\circ$ is the Schur-Product. We would like to note here that it is also possible to calculate PAI in case of a non-diagonal $\mathbf{R}$.

2. As mentioned before, the analysis is carried out independently for every model grid point. This means that also the model equivalents $\mathbf{Y}^a$ and weight vectors $\mathbf{W}^a$ will change from one grid point to another and hence also the Kalman gain. In practice however, $\mathbf{Y}^a$ and $\mathbf{W}^a$ are not stored entirely as output data since they are not required any more after the analysis has been computed. In fact, the five-dimensional field $\mathbf{W}^a$ (with three spatial dimensions of the reduced analysis grid and two ensemble dimensions) is not stored at all, because in the KENDA setup this would take the same effort and disk space as writing out about 60 additional 3D variables. The analysis model equivalents for an observation are not stored for every grid point that is within the localisation cut-off radius of the observation, but only at the grid point that is closest to the nominal position of the observation. Also for non-local observations like satellite radiance there is a nominal position that is used for the localisation. In this study, we will demonstrate that up to the localisation length scale it is a reasonable approximation to use the available $\mathbf{Y}^a$ at the nominal observation location to compute the Kalman gain at nearby grid points. This works since the weights of the LETKF by design only vary gradually from one grid point to another and with this also the model equivalents. The errors related to this approximation could be avoided by storing the full $\mathbf{Y}^a$, or $\mathbf{W}^a$ at every model grid point, but this would require significant additional disk space and would only be feasible for short experiments. In contrast, with the approximation, the PAI diagnostic can be applied to the standard output of the operational system, namely the full analysis ensemble and the model equivalents in observation space at the nominal positions of the observations.

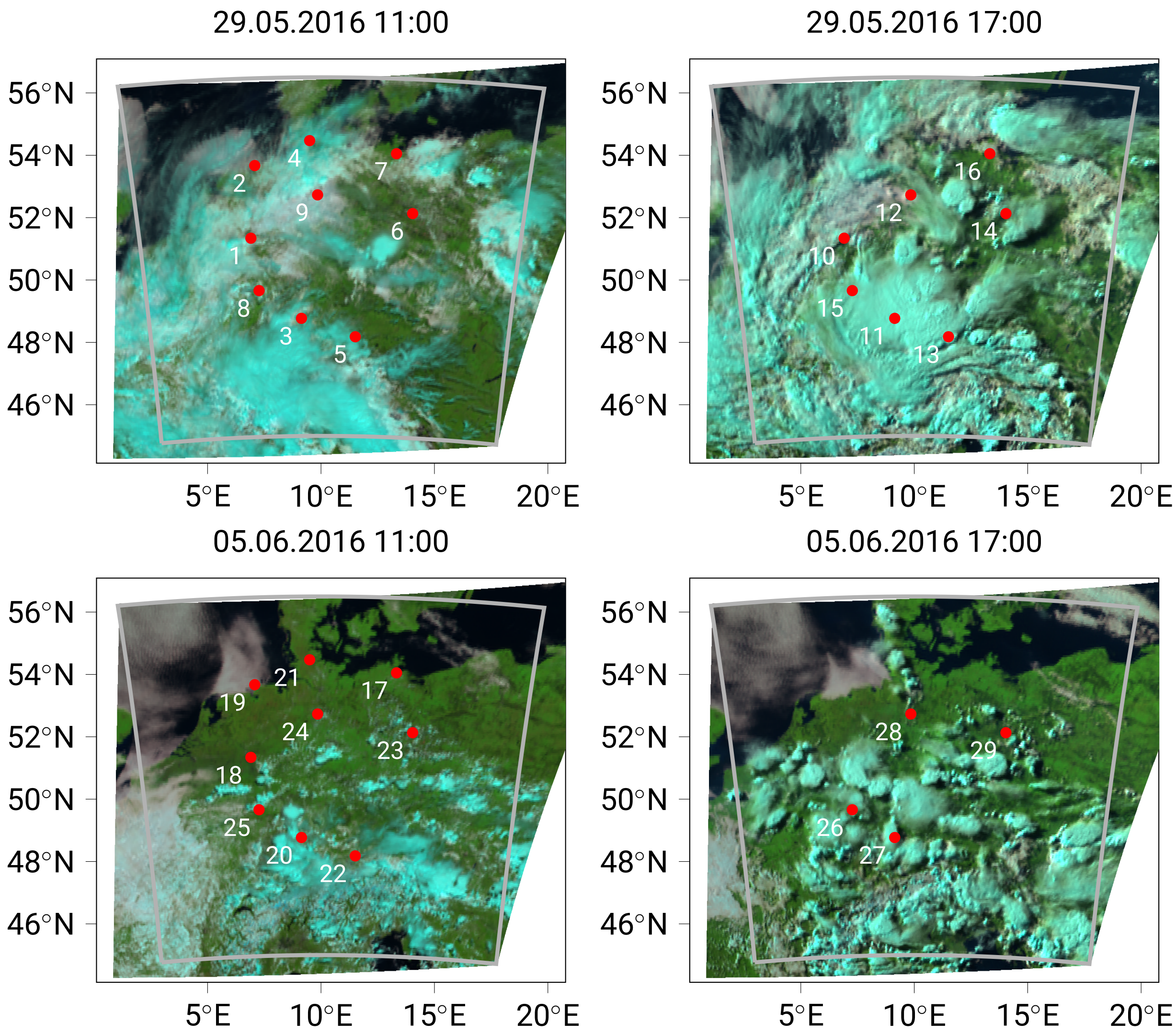
Additionally, we want to point out that the PAI diagnostic allows for computationally cheap sensitivity experiments by modifying the localisation scale or assigned observation error ($\mathbf{R}$) used to compute $\mathbf{K}$. The result yields an approximation to the influence that an observation would have with modified settings of the localisation length or the assigned observation error. This is an approximation since with a varying localisation or $\mathbf{R}$, the analysis products $\mathbf{X}^a$ and $\mathbf{Y}^a$ would also change. However, we will demonstrate that the PAI results from non-localized LETKF experiments with retrospective localisation in the PAI calculation are a useful first-order approximation for PAI in assimilation experiments with direct localisation (Section 3.3.3).

3.2.2 Description of the DA System

The configuration of the KENDA simulations used in this study closely follows that of Scheck et al. (2020). The KENDA system consists of an LETKF assimilation scheme that is coupled with the COSMO regional NWP in this study. Our experiments have 40 ensemble members and we use version 5.2 of the non-hydrostatic NWP model COSMO in its limited area configuration (COSMO-DE). COSMO was operational at DWD until April, 2021. The COSMO-DE domain is depicted in Figure 3.1 (grey box). It reaches from 44.7° to 56.5° North and from 1.0° to 19.4° East and comprises Germany and parts of its neighbouring countries. The numerical grid consists of 421 x 461 columns, resulting in a horizontal grid spacing of 2.8 km. In the vertical, COSMO has 50 hybrid layers, which are terrain-following in the lower atmosphere and flat at higher levels. The model top is at 22 km. Deep convection is resolved explicitly in the model, whereas shallow convection is parameterized. The lateral boundary conditions are interpolated from the ICON-EU model with a 7 km horizontal grid spacing and parameterized convection. For more details about the model set-up, the reader is referred to Scheck et al. (2020).

3.2.3 Experimental Set-up

To validate the methodology and to illustrate potential applications, three different types of experiments were performed: (1) Single-observation experiments for visible satellite observations, (2) single-observation experiments for visible satellite observations with collocated radiosonde observations and (3) single-observation experiments for visible satellite observations with collocated radiosonde observations using vertical localisation. This set of experiments was chosen as we primarily want to investigate the influence of novel satellite observations. Radiosonde observations are used as a reference and to investigate how the influence of the satellite observations is modified if additional observations are assimilated. The last type of experiments is used to evaluate the PAI approximation for modified assimilation settings. However, it should be noted that these observation types only serve as an example and that the PAI diagnostic is not restricted to specific observation types. For the experiments, analysis ensembles were computed for four different cases, namely 29 May 2016 11:00 UTC and 17:00 UTC and 5 June 2016


29.05.2016 11:00

29.05.2016 17:00

05.06.2016 11:00

05.06.2016 17:00

Figure 3.1: SEVIRI images at the four different dates of the experiments. The grey box indicates the COSMO-DE domain, the numbered red dots indicate the observation locations.

at 11:00 UTC and 17:00 UTC. These cases are within a highly convective period that included extreme precipitation events and a high number of consecutive severe thunderstorms over Germany. The synoptic situation in this period has been studied extensively by e.g. Keil et al. (2019), Bachmann et al. (2020) and Piper et al. (2016). For each of the cases, background forecasts were initialized at 10:00 UTC and 16:00 UTC from a reference cycling experiment where only conventional observations were assimilated (i.e. SYNOP stations, radiosondes, wind profilers, and aircrafts (AMDAR)). A single cycle with an assimilation window of one hour was computed for each of the experiments. Inflation methods were switched on in the reference cycling experiment but switched off for the last analysis step. The experimental set-up closely follows the experimental set-up of the single-observation experiments from Scheck et al. (2020) (cf. Section 3.1 and 3.2 therein).

Single-observation experiments (VIS)

In these experiments, visible satellite radiances of the 0.6 µm wavelength channel (REFL) were assimilated. The fast, look-up table based method of Scheck et al. (2016) and an approximation accounting for three-dimensional radiative transfer effects (Scheck et al., 2018) were used to generate model equivalents. The horizontal localisation length scale was set to 25 km, resulting in a cut-off radius of $\sim$ 90 km. The observation locations (shown in Figure 3.1) were chosen such that, with the given localisation length scale, the influences of the different measurements do not overlap. As we are only interested in the analysis influence, it is therefore possible to conduct multiple single-observation experiments in one model run. In total, we have 29 single-observations experiments distributed over the four time points. In contrast to thermal infrared channels, the visible channel considered here is sensitive to clouds at all heights and there is no peak in the weighting function that could be used for vertical localisation. Therefore, no vertical localisation was applied and the nominal height of all satellite observations was set to 500 hPa (following Scheck et al., 2020). Consequently, each of the satellite observations influences the whole atmospheric column within the horizontal cut-off radius. Only individual satellite pixels were assimilated. The resulting analysis departures are verified against spatio-temporally close radiosonde observations that are not actively assimilated in this experiment. This experiment is used to verify the PAI diagnostic by comparing the computed partial increments with the increment obtained from the LETKF ($x^a - x^b$). Apart from the approximations in the PAI diagnostic due to localisation both increments should be identical.

Combined experiments (RASO+VIS)

These experiments use the same setup as the ones described above, but with the assimilation of additional nearby radiosonde observations. That means we have 29 radiosonde profiles that are assimilated with a localisation length scale of 25 km in the horizontal and with a constant vertical localisation length scale of 0.3 in logarithmic pressure coordinates, i.e. log(p) = 0.3 where p is the pressure in Pa. Each of the profiles consists of $\sim$ 30 measurements of temperature (T), horizontal wind (U, V) and relative humidity (RH) distributed at different heights. This experiment shows how the influence of the satellite observations changes if additional observations are assimilated.

Combined and localized experiments (RASO+VISLOC)

These experiments use the same experimental set-up as the combined experiment, but the satellite observations are localized in the vertical using the Gaspari-Cohn function with a constant localisation length scale of 0.3 (in logarithmic pressure coordinates). The nominal positions of all satellite observations in the experiment are set to $p = 500\ hPa$. This experiment is used to investigate the feasibility of retrospective localisation of the RASO+VIS experiment in the PAI diagnostic.

Metrics and notation

For the experimental evaluation we use three different metrics: (1) the differences between the computed PAIs and the increments as obtained from the LETKF, (2) statistics of PAIs (mean, standard deviation and absolute mean) and (3) errors of the model state with respect to the radiosonde measurements, i.e. negative background and analysis departures or in case of Figure 3.9 and 3.12 absolute values of departures. We will assign the following sub- and superscripts to PAI in order to specify it correctly:

$$PAI_{exp}^{y \rightarrow x}, \tag{3.16}$$

where y represents the measured variable, i.e. $y \in$ [T, U, V, RH, REFL] (REFL is the observed visible satellite radiance and the other variables are the radiosonde measurements), x stands for the model variable for which the PAI is computed and $exp \in$ [VIS, RASO+VIS, RASO+VISLOC] indicates the associated experiment.

We evaluate the error e of the model state based on the absolute value of the difference between independent radiosonde observations and model equivalents:

$$\boldsymbol{e}_v = |H(\boldsymbol{x}^v) - \boldsymbol{y}^o|, \tag{3.17}$$

where $v \in$ [a, b] indicates whether the deviation from the radiosonde measurement is computed from the background or the analysis. It should be noted that the difference in Eq. (3.17) also contains a contribution from the radiosonde observation error. But as the radiosonde observation error is the same for the background and analysis departure and usually uncorrelated with model error, the error reduction by data assimilation can be approximated as:

$$\Delta \boldsymbol{e} = \boldsymbol{e}_a - \boldsymbol{e}_b. \tag{3.18}$$

A negative $\Delta \boldsymbol{e}$ indicates a reduction of the error and hence a beneficial impact of assimilated observations. For all the experiments the results are evaluated up to 200 hPa.

For the optimisation of vertical localisation of satellite observations, we define a cost function J that consists of the sum of the radiosonde analysis departures:

$$J(\ell, p) = \sum (H(\boldsymbol{x}^a) - \boldsymbol{y}^o_{RASO})^2. \tag{3.19}$$

here the ℓ corresponds to the vertical localization length scale and the p to pressure around which the Gaspari-Cohn function is centered. With Eq. (3.6) this can be expanded to,

$$J(\ell, p) = \sum H(\boldsymbol{x}^b + \rho(\ell, p) \cdot \mathbf{K}(\boldsymbol{y}^o - \boldsymbol{y}^b)) - \boldsymbol{y}^o_{RASO})^2 \tag{3.20}$$

More details about the optimisation are provided in Section 3.4.3.

3.3 Illustration of Partial Analysis Increments

Throughout this study, we will illustrate the PAI diagnostic exemplarily with temperature increments as the main characteristics of the diagnostic are similar for all model variables. Increments on variables other than the temperature are only shown in Section 3.4.1, where we demonstrate how to analyse the influence of observations on different model variables.

3.3.1 The Effect of Approximating PAI with Analysis Perturbations

Figure 3.2 shows the horizontal analysis temperature increment at model level 23, corresponding to a mean pressure of around 500 hPa from a single-observation experiment (VIS) that assimilated one satellite reflectance observation in the centre of the domain (Figure 3.2b) and the corresponding PAI of this observation (Figure 3.2a). The comparison demonstrates that the PAI calculation is able to reproduce both the structure and the magnitude of the analysis increment with the exception of small differences at larger distances, close to the localisation cut-off radius. These small differences are due to the approximation described in Section 2: Instead of the LETKF weights at every model grid point, the PAI calculation is based on the weights at the point of the observation expressed by the analysis perturbations to avoid the need of storing additional quantities and for the sake of computational efficiency. In the presence of localisation, the LETKF weights gradually change from one grid point to the next one leading to a deviation of PAI from the analysis increment with increasing distance from the observations. This difference can also be seen in the comparison of the PAI and the analysis difference as a function of horizontal distance from the observation in Figure 3.2d. However, the LETKF weights by design only change very gradually from one grid point to the next one. This means that the differences of the efficiently approximated PAI and the analysis increment are fairly small and avoiding these small differences doesn't seem to justify the additional storage of LETKF weights. In the vertical, the calculated PAI perfectly matches the analysis increment (except very small rounding errors) as no vertical localisation was used for the assimilation of the satellite observation (Figure 3.2c).

Besides the example shown in Figure 3.2, we calculated further single-observation experiments for 29 reflectance observations. Figure 3.3 shows a comparison between the absolute analysis increment and the absolute difference of the analysis increment minus the computed PAI as a function of distance from the observation for all single-observation experiments. On average, the difference of PAI and the analysis increment is less than 17% up to the localisation length scale of 25 km and increases to about 40% at twice the localisation length scale.

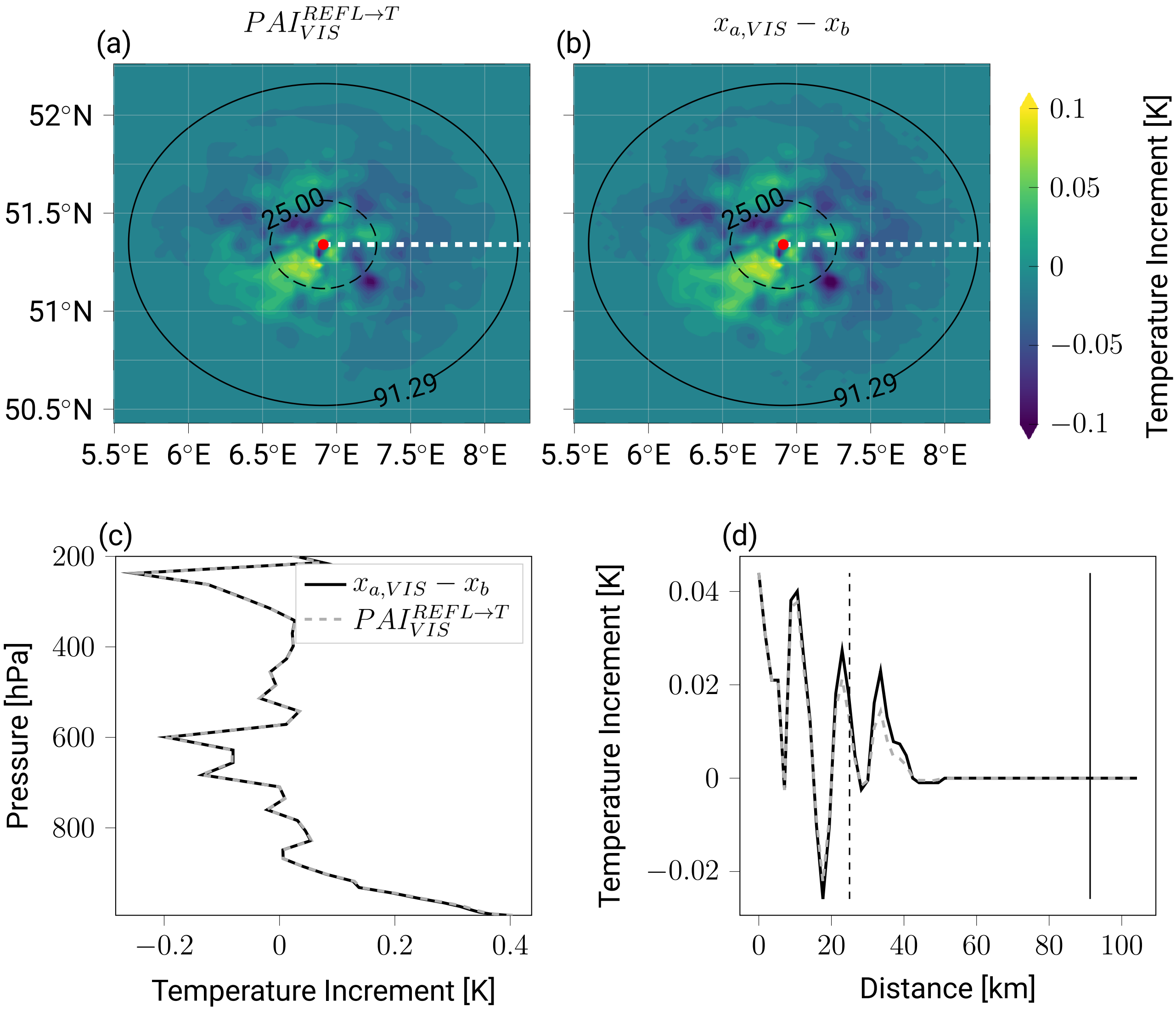


Figure 3.2: Temperature increments for one VIS-Experiment (at location Number 1 in Fig. 3.1). (a) The computed PAI at one model level at $\sim$ 500hPa. The shading indicates the magnitude of the temperature increment, the dashed circle indicates localisation length scale and the solid circle the cut-off radius. (b) same as (a) but shows the increment as obtained from the LETKF run. (c) Vertical profile of the increments at the observation location (red dot in the upper panels). (d) Increments as a function of horizontal distance from the observation, horizontal cut through the domain along the white dashed line in the upper panels, the dashed and solid vertical lines indicate localisation length scale and cut-off radius, respectively.

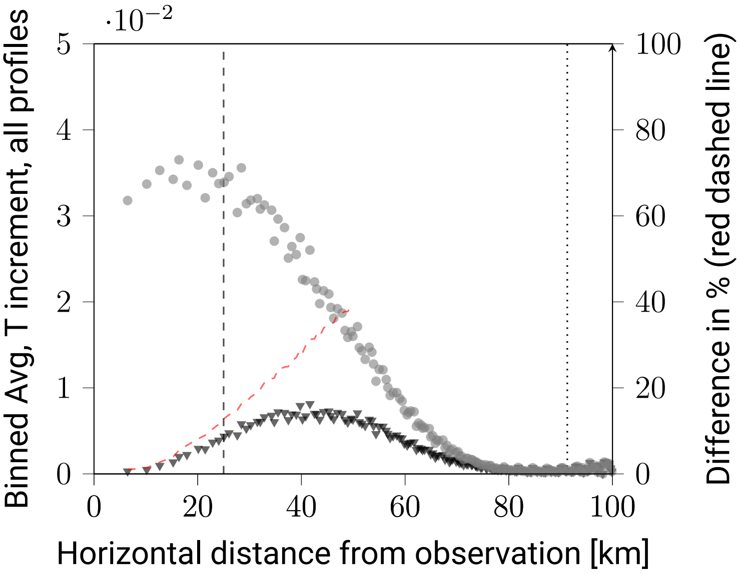


Figure 3.3: Binned averages of the temperature increments as a function of horizontal distance from the observation. The averages are taken over all 29 observation locations in the VIS-Experiment at one model level at $\sim$ 500 hPa. (grey dots) The absolute analysis increment as it is obtained from the LETKF run ($x_{a,VIS} - x_b$). (black triangles) The absolute difference between the LETKF analysis increments and the computed PAIs. (red dashed line) The relative difference in % between PAI and the LETKF increment. The black dashed line indicates the localisation length scale, the dotted line indicates the cut-off radius. Values to the right of the cut-off radius come from neighbouring single-observation experiments. Deviations between the computed PAI and the LETKF increment are due to approximations in the PAI diagnostic.

3.3.2 The Relation of PAI with the Increment from Single-Observation Experiments

The PAI of an observation from an experiment assimilating many other observations is by nature not the same as the respective analysis increment in a single-observation experiment. The presence of other assimilated observations decreases the weight of an individual observation compared to the single-observation experiment. In case of multiple observations with equal observation errors assimilated at the same location, the weight of every assimilated observation will decrease by the following factor with the addition of a further observation:

$$\frac{\alpha^{n+1}}{\alpha^{n}} = \frac{n}{n+1} \tag{3.21}$$

where α^{n+1} is the weight of observations in case of $n+1$ observations and α^{n} is the weight of observations in case of n observations. Figure 3.4 shows this effect for up to 40 assimilated observations. The number of 40 reflects the local degrees of freedom of the 40-member LETKF system and therefore the order of magnitude of observations that can be assimilated within the localisation scale. Assimilating two observations instead of one decreases the weight by a factor of 0.67. With more assimilated observations, the factor gradually increases to 0.91 for 10 observations and 0.98 for 40 observations. This means that adding an additional observation in a comprehensive data assimilation with many assimilated observations only has a marginal effect on the weight of the other assimilated observations.

It is important to keep this effect of modified weights in mind when interpreting PAI results. However, the addition of other observations only reduces the weight of an observation, but does not change the overall structure of the influence of an observation. Figure 3.5a shows an example of the temperature PAI of a satellite reflectance observation in an experiment assimilating this observation and additionally a full radiosonde profile (two wind components, temperature and humidity at 39 levels; i.e. 156 additional observations, RASO+VIS) and the analysis increment in a single-observation experiment with only the satellite observation (VIS). As expected, PAI is smaller than the increment in the single-observation experiment, but both exhibit a similar structure.

Figure 3.5b shows the corresponding mean absolute PAI and single-observation analysis increment averaged over all 29 assimilated satellite observations. On average, the magnitude of the single-observation analysis increment is roughly 30-50% higher than the corresponding PAI in the experiments with radiosonde observations in addition. The structure of the profile is however very similar with largest values of increments and PAIs in the lowest and highest part of the profile. This near-linear behaviour of the influence demonstrates that both PAI and single-observation experiments are useful approaches to investigate the 3D influence of observations. The calculation of PAI, however, is computationally much more efficient. And furthermore PAI reflects the influence in the presence of other assimilated observations, which is usually the primary quantity of interest, whereas single-observation experiments reveal the influence in the absence of other observations.

Figure 3.6 shows an example of the contribution of different observations to the temperature increment (temperature PAI) as a function of pressure in the RASO+VIS experiment. As expected, radiosonde temperature observations exhibit the largest temperature PAI throughout most of the atmosphere. The satellite observation, however, also leads to a significant temperature increment in the boundary layer, which is likely related to the correlation of cloudiness and surface insolation.

Information such as the relative magnitude of increments and the strength of the downweighting effect through the assimilation of other variables cannot be retrieved from single-observation experiments alone. A statistical analysis of the PAI estimated increments on different variables will be discussed further in the next section in the context of potential applications of the diagnostic.

3.4 Potential Applications

3.4.1 Analysing the Influence of Observations on Different Model Variables

The PAI diagnostic allows for analysing the influence of individual observations as well as the statistical contribution of observation types to changes in different variables. Especially with regard to operational data assimilation, general information about the relative magnitude of increments is useful to evaluate the effectiveness of the assimilation. Moreover, statistics of PAI can be used to analyse trends (systematic increments) introduced by certain observation types. In particular for novel observations such as e.g. satellite reflectance, it is important to monitor that the observations do not cause systematic changes in the model climatology as cooling/warming or drying/wettening at

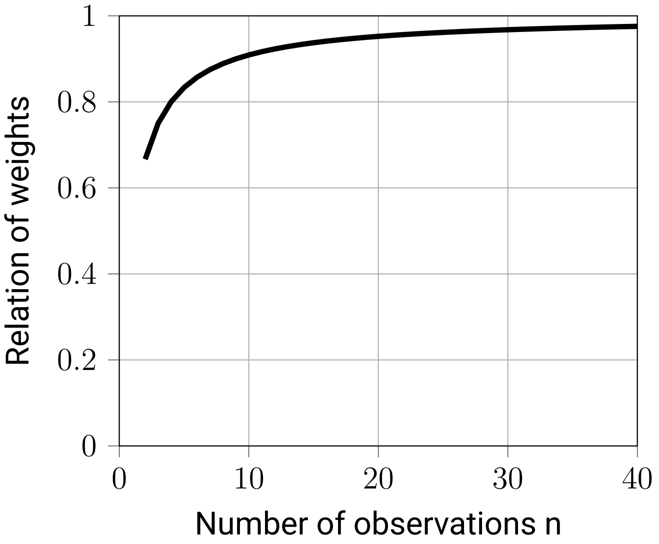


Figure 3.4: Relative decrease of weights for assimilating one additional observation as function of the number of the other assimilated observations with the assumption that all observations are at the same location and have the same observation error.

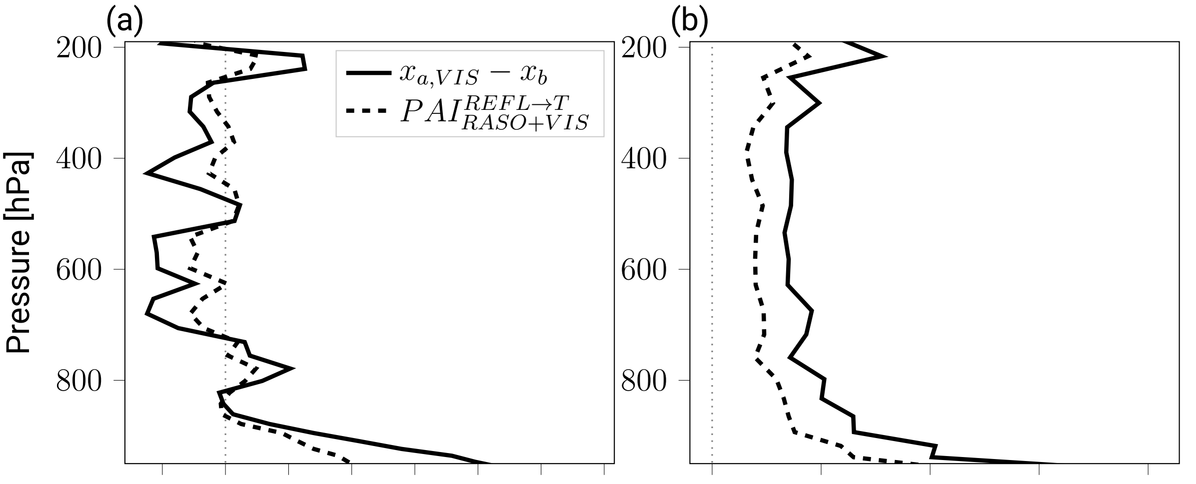


Figure 3.5: (a) Vertical profile of the temperature increment at a single-observation location (No. 20 in Figure 3.1). (b) Vertical profile of the mean absolute temperature increment at the observation location, the mean is taken over all profiles from all 29 observation locations, 30 bins in the vertical. (solid) Increment for the VIS-Experiment, where only satellite observations were assimilated. (dashed) PAI of the satellite in the RASO+VIS-Experiment.

certain levels. While similar information can be gained from single-observation experiments or EFSO, the PAI diagnostic can be considered as either an economical alternative or an economical addition to such measures which is also capable to identify systematic non-local effects as e.g. the systematic influence of the satellite observations on various vertical levels. The statistical analysis about the performed RASO+VIS experiments is shown in Table 1 and Figure 3.7. Similar as in Scheck et al. (2020) this analysis shows that in general the data assimlation of the visible satellite observations yields results with physically plausible interpretations.

Averaged over all assimilated satellite observations, the temperature PAI of satellite observations is about 5% of the total temperature increment above 750 hPa and increases gradually below to about 14% at the lowermost level (Figure 3.7a). The relative contribution of the satellite to the wind increment is overall of a similar magnitude and structure as for the temperature increment, but with a less pronounced maximum at lower levels. In absolute terms however, the satellite wind PAI is highest at upper levels given increasing wind speed with height (dashed line in Figure 3.7c). For relative humidity, the satellite also contributes to about 5% of the total increment above 550 hPa, but to 10-15% of the total increment below 550 hPa (Figure 3.7a). As humidity only has a marginal effect on satellite reflectance in the visible range, the humidity PAI of the satellite observations is likely the result of correlations of cloudiness with humidity at the level of the cloud and beneath.

Averaged over the vertical profile, the satellite observations contribute about 7% of the total increment in relative humidity and roughly 5% of the temperature and wind increment (Table 3.1). This is remarkable given that only 0.9% of all assimilated observations are from the satellite and neither wind, temperature or humidity have a pronounced direct influence on satellite reflectance in the visible range. Whether these increments also pull the analysis in the right direction will be investigated further in the subsequent section. The largest relative PAIs of satellite observations are found for cloud water (13.9%) and cloud ice contents (8.7%), which directly influence reflectance in the solar channels. Furthermore, a comparably large PAI of the satellite observations of 7.6% occurs for vertical velocity that is linked to convection and thereby to convective clouds. For radiosonde observations, Table 1 shows that direct observations of wind components, temperature and relative humidity contribute about 60% to the increment of the respective variable. The relative PAI of an observed variable on other variables is in the range of 10-15%.

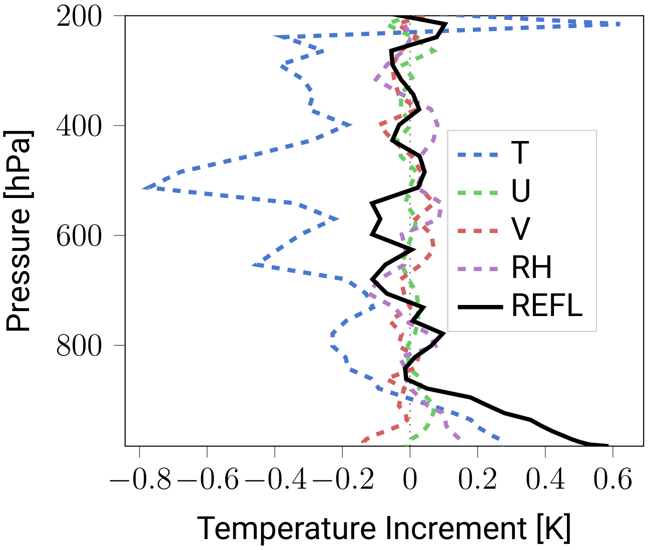


Figure 3.6: Vertical profile of different partial temperature increments from all measured variables in the RASO+VIS-Experiment for one profile (No. 20 in Figure 3.1). The sum of all partial increments of different observations equals the total temperature increment.

Model Variable \ Observation	T	U	V	RH	REFL
T	65.5	9.2	11.0	9.6	4.7
U	13.1	58.1	14.6	9.2	5.0
V	14.1	12.9	59.5	9.2	4.3
RH	12.5	9.4	11.5	59.5	7.1
WZ	28.3	19.6	28.7	15.8	7.6
Q	52.0	14.4	11.7	13.6	8.3
QI	28.5	24.2	21.4	17.2	8.7
CLWMR	23.2	15.8	29.4	17.7	13.9

Table 3.1: Relative absolute PAI contributions in % for all assimilated observations averaged over all profiles in the RASO+VIS-Experiment for the model variables temperature (T), zonal wind (U), meridional wind (V), relative humidity (RH), vertical velocity (WZ), specific humidity (Q), cloud ice (QI) and cloud water mixing ration (CLWMR). The normalization is done with respect to the total absolute increments ($x_{a,RASO+VIS} - x_b$) of the the respective model variables.

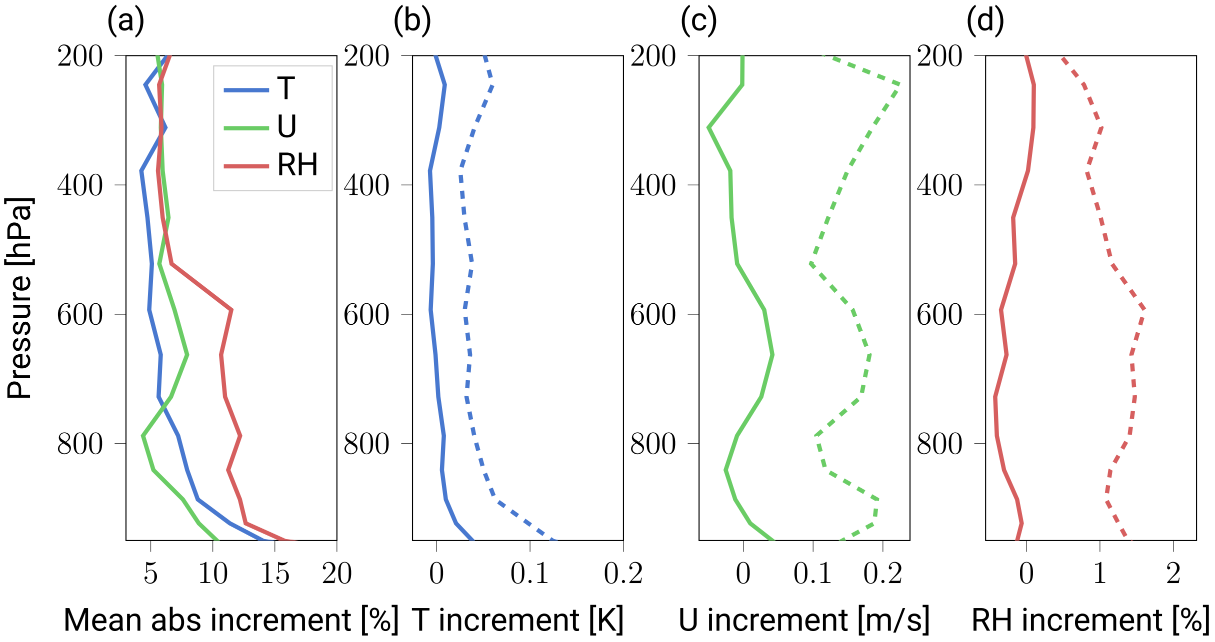


Figure 3.7: (a) Vertical profiles of the relative absolute PAI contributions in % for the assimilated satellite observations averaged over all 29 profiles in the RASO+VIS-Experiment for model variables T, U and RH. The normalization is done with respect to the total absolute increments ($x_{a,RASO+VIS} - x_b$) of the the respective model variables. (b - d) Vertical profiles for different model variable T, U, RH but this time the solid lines indicate the mean satellite PAI reflecting systematic effects and the dashed lines indicate the standard deviations of the respectives satellite PAIs reflecting their magnitudes.

3.4.2 Detecting Detrimental Observation Influence

For the assimilation of novel observation types, it is important to investigate if the assimilation of such new observations has beneficial or detrimental influence on the model state. In this study we verify the first guess and analysis states against the observed radiosonde profiles as described in Section 3.2.3. While detrimental analysis increments are part of the statistical nature of the data assimilation (Gelaro et al., 2010), extended or systematic patterns of detrimental influence indicate potential flaws in the data assimilation system and may provide guidance for optimising assimilation settings, e.g. assigned observation error or localisation parameters.

As the influence of individual observations is often blurred in cycled experiments with many observations, previous studies used single-observation experiments that assimilated only few observations separated by sufficiently large distance to avoid an interaction of the observations (e.g. Schomburg et al. (2015), Scheck et al. (2020)). In their study, Scheck et al. (2020) conclude from single observation experiments that the assimilation of visible satellite reflectance is able to reduce errors in the model state in their selected cases but that the effectiveness of this process is limited due to ambiguity of the observations, spurious correlations or non-linearity of the observations operator. In this section, we demonstrate that similar information can be gained by the PAI diagnostic and the considerable effort for carrying out additional single-observation experiments can be avoided. For this, we present PAI results as well as the analysis increments of single-observation experiments for two cases. Case 1 (profile 20 in Figure 3.1) corresponds to the same single-observation experiment as Case 1 in Scheck et al. (2020). Our Case 2, corresponds to profile 13 in Figure 3.1 is not the same as Case 2 as in Scheck et al. (2020).

Figure 3.8 shows the estimated error of background and analysis model states with respect to the radiosonde observations for the two cases. In each of the two panels of Figure 3.8 the blue line indicates the error of the background model state with respect to the radiosonde observations (negative background departure). The red line indicates the error of the analysis model state obtained in the single-observation experiment (VIS) with respect to the RASO measurement and the green line shows the negative background departure plus the satellite PAI from the RASO+VIS-Experiment. The sum of the background departure and the satellite PAI reflects the approximated contribution from the satellite to the analysis departure in the RASO+VIS-Experiment.

In Case 1, shown in Figure 3.8a, the analysis is at nearly all levels closer to the radiosonde observation than the background indicating a beneficial influence of the satellite observation in the single-observation experiment. Similar information can be gained by the computationally much cheaper PAI diagnostic that does not require additional experiments. The satellite PAI is usually smaller than the single-observation increment given the presence of other assimilated observations. However, the satellite PAI nearly always points into the same direction as the analysis increment and also indicates a beneficial influence of the satellite throughout this vertical profile. In Case 2, shown in Figure 3.8b, both the PAI and the single-observation experiment indicate a beneficial influence of the satellite observation around 900 and 300 hPa, whereas there is indication for deterioration at 240 hPa.

Figure 3.9 shows a scatter plot comparing the computed impact of the satellite measurements on the model state (Δe) in the combined experiment (RASO+VIS, y-axis) and in the single-observation experiment (VIS, x-axis) for all assimilation experiments at all radiosonde observation levels. The results of the satellite impact in the RASO+VIS-Experiment were obtained from the PAI diagnostic. Negative values of Δe indicate that the satellite observation draws the model temperature closer to the radiosonde observation (beneficial impact) and positive values indicate detrimental impact. Overall, there is a clear correlation of beneficial and detrimental impacts from the two approaches. The slope of the linear fit is close to 0.5 indicating that the impact in the single-observation experiment is about twice as large. Most importantly, both approaches indicated the most beneficial and most detrimental impact at the same locations. The largest beneficial impact occurs at profile 20 at low levels and profile 29 in the mid-troposphere. The largest detrimental values occur at upper levels for profiles 13, 19, and 20 as well as at low levels for profiles 22, 25, and 29. As in previous studies the results of Figure 3.9 show that there is a large number of observations with detrimental influence on the analysis. This is on the one hand related to the analysis verification with radiosonde observations and on the other hand related to the statistical nature of the data assimilation system. Addtionally, we want to mention that this application has a lot in common with 0h-EFSO from (Hotta et al., 2017a), as 0h-EFSO reflects the partial analysis increment projected onto a specific norm (e.g. total energy). This illustrates that the PAI diagnostic can be used to identify potential detrimental effects that should be investigated in more detail with other diagnostics or by additionally approximating assimilation settings as e.g. the localisation scale with the PAI diagnostic. The latter option will be discussed further in the subsequent section.

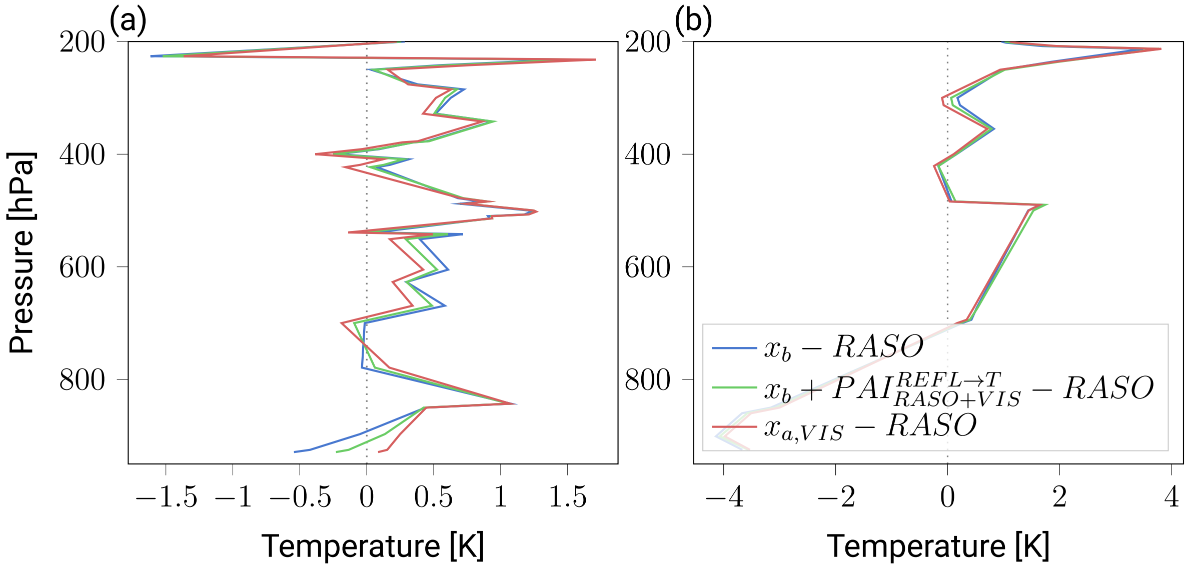


Figure 3.8: Vertical profiles of the estimated error of the model state with respect to the radiosonde measurements. (a) Profile of Case 1 with highest error reduction in the boundary layer. (b) Profile of Case 2, minor corrections but also some deterioration in the upper atmosphere. In both panels: (blue) background minus observation (green) analysis minus observation computed with $PAI^{REFL \to T}_{RASO+VIS}$ from the RASO+VIS-Experiment (red) analysis minus observation from the VIS-Experiment.

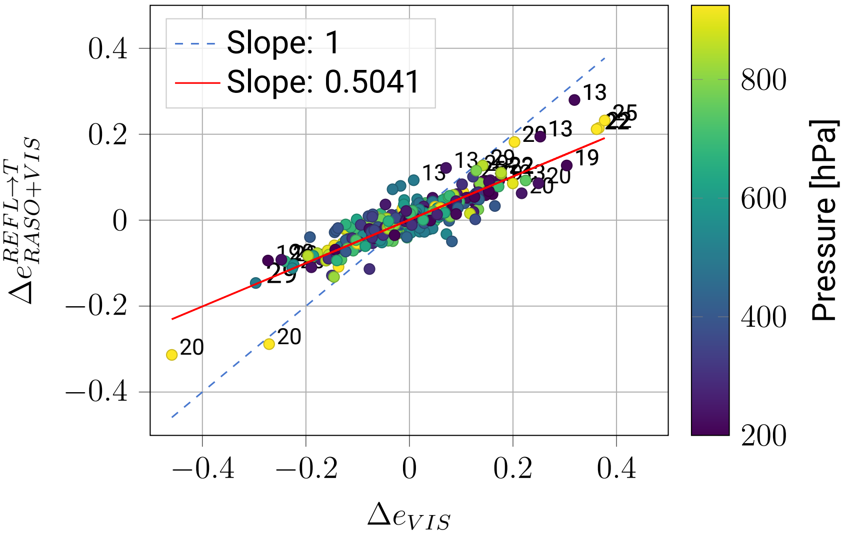


Figure 3.9: Scatter plot comparing the impact of the satellite measurements on the temperature in the single-observation experiment (Δe_{VIS}, x-axis) versus in the combined experiment ($\Delta e^{REFL\to T}_{RASO+VIS}$, y-axis). The impact is measured by the change in the temperature errors due the assimilated reflectances. The dots indicate all radiosonde observation levels of all profiles. The color shading indicates the pressure level. The errors are measured with respect to the radiosonde observations. The number labels indicate the profile number as shown in Figure 3.1.

3.4.3 Optimising Localisation

In the last section, we discussed that PAIs can be used to detect detrimental observation influence caused by suboptimal assimilation settings. Additionally, the PAI diagnostic can be used to approximate the influence of observations assimilated with modified settings for localisation or the assigned observation error without rerunning the assimilation cycle. To demonstrate this, we retrospectively localised the satellite PAI from the RASO+VIS-Experiments vertically with a localisation scale of 0.3 centered at 500 hPa and conducted assimilation experiments with a corresponding localisation for satellite observations (RASO+VISLOC). Figure 3.10 shows that vertical localisation strongly reduces the influence of satellite observations at lower and upper levels as expected. Furthermore, Figure 3.10 demonstrates that PAI with retrospective localisation (red line) is a good approximation of PAI in the RASO+VISLOC experiment with localisation for satellite observations (green line). Only minor differences occur between the retrospective localisation in the PAI calculation and the localisation in the assimilation system.

This means that with the retrospective vertical localisation, it is possible to approximate optimal localisation settings in a computationally cheap manner. The concept is to define a cost function based on the analysis departures of observations that are not assimilated and minimise this function iteratively with respect to the localisation settings. The cost function J is defined in Eq. (3.19). In our study, we used the analysis departures of passive radiosondes and the satellite PAIs computed in the VIS-Experiment to demonstrate the concept. For localisation with the Gaspari-Cohn function, the localisation length scale and the height at which the Gaspari-Cohn function is centered can be optimised. In Figure 3.11 the cost function was computed for all profiles in the VIS-Experiment. For the iterative optimisation with respect to localisation length scale ℓ and cloud height p, we find that the optimal $\ell = 0.4$ and the optimal $p = 800\ hPa$. As compared to no vertical localisation of the satellite the optimal localisation with the Gaspari-Cohn function improves the analysis departure statistics by 1.5 % (red dot in Figure 3.11).

As Figure 3.9, Figure 3.12 shows how the analysis increment from the satellite single-observation experiments and the corresponding satellite PAIs draw towards the radiosonde observations. The underlying light grey dots are the previously shown results without localisation and the coloured dots are the results with retrospective vertical localisation with the computed optimal localisation settings. Detrimental effects mainly occurred in the boundary layer and at high levels without localisation. The localisation reduces large positive values in the upper atmospheric layers. The largest negative (beneficial values), which are linked to increments in the boundary layer, are only slightly modified. This illustrates that the PAI diagnostic can be used for efficiently testing various localisation approaches without rerunning the assimilation experiments. However, it should be noted that the optimised satellite localisation in this study was derived from a small sample size for illustrating the concept. Deriving general conclusions for the localisation of satellite reflectance will require longer experiments that are planned for future studies. Furthermore, it should be noted that in case of the VIS-Experiment, PAIs are equal to the respective total analysis increments as no other observations are assimilated and the observations are at distances larger than the horizontal localization radius. Nevertheless, we illustrated this approach as it would equally be applicable in an experiment assimilating the full observing system, where PAIs would identify the individual influence of individual observations and thereby serve as basis for optimising localisation. For our example, we also tested to construct a cost function based on assimilated radiosondes in the RASO+VIS-Experiment, but achieved no meaningful results. Thus, we think that independent (passive) observations are required for optimising localisation. The implementation of this approach in a near-operation DA system may also need to account for specific system settings, such as adaptive inflation or localization. Moreover, we expect that the results of the optimization also depend on the region and the synoptic situation that is considered. In contrast to our experiments covariance inflation is typically used in a near-operational set-up to counter overconfidence of the analysis and give more weights to the observations. The implementation of inflation in the PAI computation depends on the inflation technique that is used, e.g. prior or posterior inflation.

3.5 Conclusions

This study proposes to use partial analysis increments (PAIs) as diagnostic for LETKF data assimilation systems. The exact computation of these increments would require large amounts of additional output from the LETKF in the form of the five-dimensional weight matrix that is not available in operational setups. However, the results presented here demonstrate that PAI can be approximated efficiently using ensemble analysis perturbations available from the standard LETKF output. We demonstrate that using analysis perturbations instead of ensemble weights only introduces very minor errors at larger distances from the observations. Furthermore, we analyse the difference of observation influence in single-observation experiments with cloud-affected satellite observations

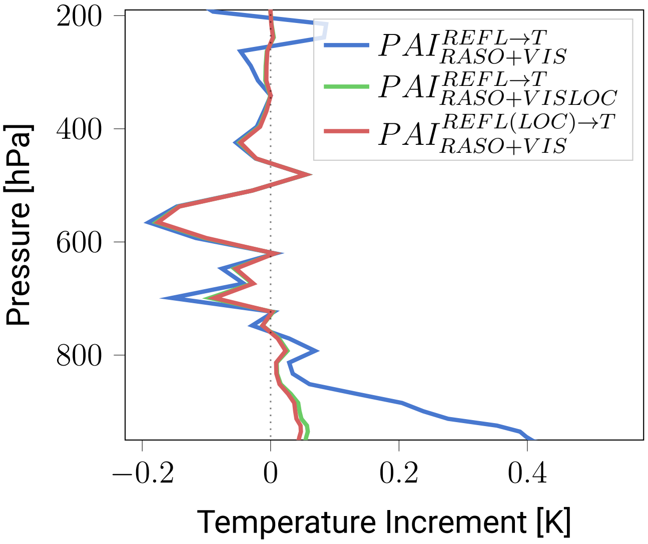


Figure 3.10: (blue line) Vertical profile of satellite temperature PAI from the RASO+VIS-Experiment without localisation for satellite observations, (red line) corresponding PAI profile from the experiment RASO+VISLOC with localisation for satellite observations and (green line) corresponding profile of satellite PAI with retrospective localisation in the PAI calculation from the RASO+VIS-Experiment.

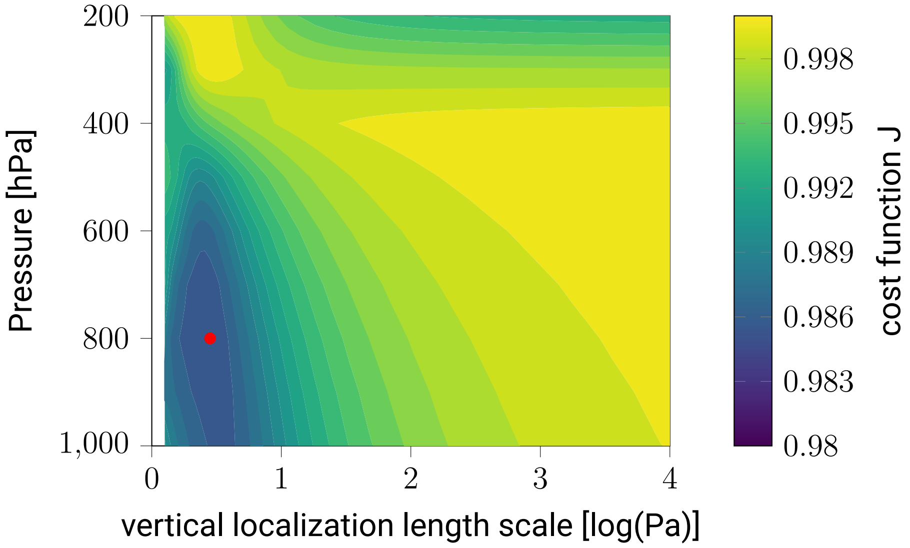


Figure 3.11: Contour plot of the cost function J as defined in Eq. 3.19 as a function of localization length scale (x-axis) and centering height of the Gaspari-Cohn function (y-axis). J is computed iteratively with retrospective localization. The J computed with no vertical localization is set to 1.

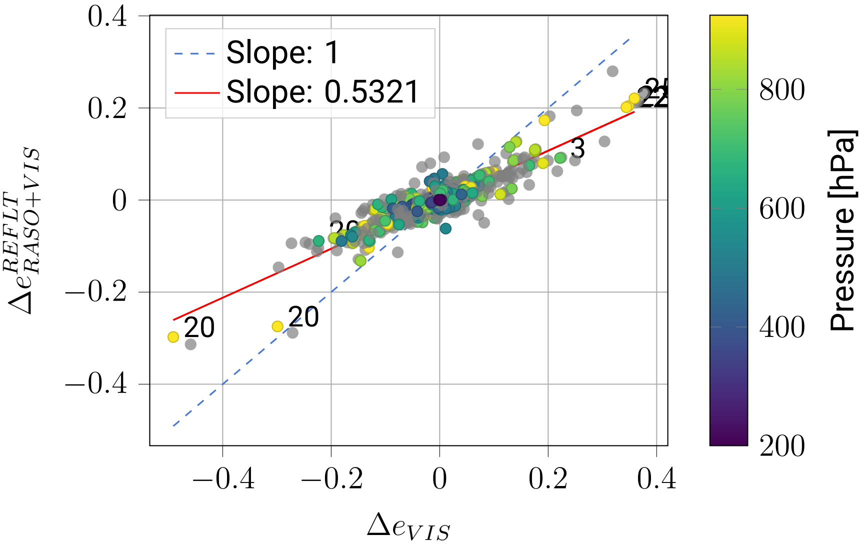


Figure 3.12: Same as Figure 3.9 but the colored dots show the change in temperature errors due to the retrospectively localised satellite PAIs with the computed optimal localisation settings for the Gaspari-Cohn function. The grey dots are the same as in the previous figure for the non-localised satellite PAIs from the RASO+VIS-Experiment.

in the visible spectrum and PAI in experiments that assimilate both radiosondes and satellite observations. The influence of an observation is decreased by the presence of other assimilated observations, but we demonstrate that this effect primarily leads to a reduced influence and does not change the structure of the influence significantly. This means that both single-observation experiments and PAI can be used to investigate the influence of promising additional observations as e.g. satellite radiances. The PAI approach, however, is computationally much more efficient and has the advantage that it directly reflects the influence of observations in the presence of other assimilated observations which is usually the primary quantity of interest.

Additionally, the study illustrates and discusses three potential applications of PAIs as a diagnostic method. First we show that PAI can be used to analyse the contribution of different observations to the analysis. In contrast to other scalar diagnostics for observation influence, PAI describes the full 3D influence on the analysis state. This means that also non-local effects of observations can be analysed as well as their effect on other variables besides the observed quantity. We illustrate this approach based on experiments that assimilated experimental satellite observations and radiosondes, where it can be seen that satellite observations also contribute to e.g. model temperature, in particular in the atmospheric boundary layer. Besides the use of the diagnostic for investigating the detailed effects of novel experimental observations shown here, the diagnostic also appears valuable for monitoring more complex operational assimilation systems with multiple observations types. In contrast to a monitoring based on departures and increments in observations space, this would also allow to detect e.g. non-local trends introduced by some observation types (e.g. systematic drying/wetting in some regions). The PAI diagnostic therefore offers a computationally inexpensive approach for monitoring and analysing operational data assimilation systems.

Secondly, we show that PAI can be used to detect where different observations draw the analysis in opposite directions as indicator for suboptimal assimilation settings or erroneous observations. The approach is validated with single-observation experiments that show good overall agreement with the PAI diagnostic. Our study primarily focuses on the effect of the experimental satellite observations and determines, where their influence is in the same or the opposite direction to radiosondes. The same approach, however, could be used in an operational system to automatically detect large discrepancies between the influence of different observations or observation types.

Last but not least, we show that PAI can also be used to approximate the influence that observations would have with modified assimilation settings with the example of a modified vertical localisation scale for the satellite observations. This approach includes a second approximation which is the modification of the influence of other observations. The comparison to additional experiments with modified localisation scale, however, again shows that the approximation only has a comparatively minor effect. For the conducted experiments, we show that vertical localisation removes the largest opposing influence of satellite and radiosondes observations that is likely due to spurious ensemble covariances. However, this comes at the cost of also removing beneficial (corresponding) influences in some regions. How to optimally treat vertical localisation for cloud-affected satellite observations is subject of other ongoing research projects, but the PAI diagnostic provides an efficient tool to investigate various potential approaches without the need for additional experiments for every configuration. Furthermore, it could be used to objectively optimise the localisation length scale based on the minimisation of opposing influences in a larger data set similar to the approach of Hotta et al. (2017b) for optimising the observation error covariance matrix. We did not discuss covariance inflation although it is another major tuning parameter in data assimilation systems and should be subject to further research in this context. We expect that the computation of PAI can be extended to take into account inflation, the details, however, will depend on the inflation technique that is used in the DA system. In principle PAI can indicate cases where observations have very small influence and might therefore also give indication of regions with too little ensemble spread. Hence, the PAI diagnostic could provide a basis also for the investigation of adaptive inflation methods.

Chapter 4

Diagnostics for Imbalance on the Convective Scale

Data assimilation has the potential to introduce imbalance, which may degrade the quality of the forecast. After having discussed, predominant balance principles on different scales in the atmosphere in Section 2.2, the conclusion has been drawn that imbalance on the convective-scale is elusive and even the question how to measure the imbalance on the convective scale is unsolved. In this chapter, we compare three measures of imbalance that are potentially relevant to convective scale data assimilation. The measures are based on different physical arguments, thereby addressing Research Question 2: *Which measures of imbalance are suitable for the convective scale?*
This chapter is based on: Diefenbach, T., Scheck, L. & Weissmann, M., Craig, G.(2024) Diagnostics for Imbalance on the Convective Scale, submitted to Monthly Weather Review

4.1 Measures of Imbalance

The first diagnostic is the surface pressure tendency, which is probably the most commonly-used measure of imbalance. In the hydrostatic approximation the rate of change of surface pressure is equal to the column-integrated mass divergence. It is thus related to fast waves that project on to surface pressure, such as the external gravity wave mode, and can be thought of as measuring one aspect of imbalance, although other aspects, such as some internal gravity wave modes, will not be seen. Surface pressure tendency has proven very useful in practice, since its time evolution during numerical forecasts consistently shows an initial transient behaviour that decays. This is true in km-scale simulations, as well as in coarser-resolution models where the hydrostatic approximation is more obviously valid.

The second imbalance diagnostic is the vertical velocity variance in lower troposphere, in the vicinity of convective clouds. This quantity was found by Lange et al. (2017) to be directly connected with triggering of spurious convective cells. To focus on gravity waves that are responsible for triggering spurious cells, Lange et al. (2017) computed vertical velocity variance, masking out the interior of the convective updrafts, and also regions far from convection, where gravity waves are more likely to be associated with orography.

The third diagnostic is based on the proposal of Craig and Selz (2018) that the weak temperature gradient (WTG) approximation holds to a useful degree on the convective and mesoscales, so that the departure from WTG is a measure of imbalance. Klein (2010) shows that WTG is a consistent approximation for motions with speeds that are slow in comparison to the propagation speed of internal gravity waves. On synoptic scales, the atmosphere adjusts to a state of geostrophic balance after the transient gravity waves have propagated away (as in the Rossby adjustment problem; Bierdel et al., 2017). On meso- and convective scales, the Coriolis force is not dominant, and horizontal pressure and temperature gradients are small in the adjusted state. The balanced response to a heat source, such as latent heat release in a cloud, is therefore to produce ascent so that adiabatic cooling equals heating. In this way, no local temperature anomaly is produced by the heat source. A deviation of vertical velocity from this value indicates the presence of fast transient motions that are not part of the balanced flow. A deviation of vertical velocity from this value indicates the presence of fast transient motions that are not part of the balanced flow. If the heating rate changes only slowly, the amplitude of the transients will remain small throughout the adjustment, but this will only be partially true for convective clouds and may not be true at all for data assimilation increments. By measuring the amplitude of transient variations in the thermodynamic equation in high-resolution numerical simulations, Craig and Selz (2018) showed that the WTG approximation holds approximately for convective and mesoscale motions in the midlatitudes, although it is a less accurate approximation than geostrophic balance on the synoptic scale.

4.2 The Role of Data Assimilation Method and Convective Weather Regime

To explore the performance of the different imbalance metrics, our approach involves applying them to analyses and first-guess forecasts from diverse data assimilation experiments. These experiments utilise different DA methods, and as a result, exhibit varying levels of imbalance in the analysis, enabling us to draw conclusions about the usefulness of each metric. The focus of this study is on ensemble data assimilation and the NWP system that we use to conduct the experiments is the ICON-KENDA system (Schraff et al., 2016), developed at the German Weather Service (DWD).

The first convective-scale DA method that is considered is latent heat nudging (LHN), which relies on the assumption that the latent heat release is proportional to the rain rate at the surface as measured by radar reflectivity data (Stephan et al., 2008). LHN is implemented in the ICON-KENDA system. If the model underestimates the radar-observed rain rate a positive temperature increment is added to the modeled state, which leads to upward motion, condensation and a resulting enhanced rain rate in the model. If the model overestimates the rain rate the temperature increment is negative. A humidity increment is applied alongside the temperature increment, since otherwise the temperature increment would be offset by saturation adjustment after the LHN. Since this method, unlike more recent DA methods, disregards correlations of any kind in the model state, it is expected that the LHN will lead to considerable amounts of imbalance.

Secondly, the degree of imbalance generated by the implementation of the local ensemble transform Kalman filter (LETKF, after Hunt et al., 2007) in ICON-KENDA is investigated using the metrics. In the LETKF, the background error covariance matrix, which describes the uncertainty of the model, is estimated by means of an ensemble. Due to the limited number of ensemble members this estimation is prone to sampling error and spurious correlations. To mitigate this effect, the LETKF relies on localisation, which typically confines the spatial impact of an assimilated observation to a local region around its nominal position. Both sampling error and localisation can result in unphysical updates of the model and may introduce imbalance in the analysis state (see e.g. Mitchell et al., 2002; Greybush et al., 2011)..

Thirdly, the LETKF in combination with the incremental analysis update (IAU) will be evaluated in terms of imbalance. The IAU distributes the increment computed at the analysis time gradually over the assimilation window, and therefore acts as a kind of low-pass filter that reduces high-frequency noise. Originally, it was developed by Bloom et al. (1996) for synoptic-scale DA with assimilation windows of the order of 6 hours. Nowadays, 3D and 4D (Lei and Whitaker, 2016) versions of the IAU exist. In the 3D version, constant "drips" of the increment are added to the model state, whereas the 4D version accounts for the temporal evolution of the flow during the assimilation window with time-varying incremental updates (see e.g. Stratman and Brewster, 2017; Ge et al., 2022). The ICON-KENDA system has adopted the 3D IAU for the convective scale, which means that the computed LETKF analysis increment is added over a time frame of 10 minutes (starting 5 minutes before the analysis time). While it is known that the IAU decreases the initial peak of surface pressure tendency in the analysis, there is lack of systematic research on the effect and the optimal parameter settings of the IAU on the convective scale. In this study, we use the IAU as implemented in ICON-KENDA to investigate the reduction of imbalance as compared to pure LETKF DA.

We will apply the different DA techniques and imbalance diagnostic to two distinct weather situations. The first case is a weak forcing case, where convection is mostly triggered by smaller scale features. The second case exhibits strong synoptic forcing, where the convection is mostly determined by the large scale flow (Keil et al., 2014).

This section is structured as follows. Section 4.3 provides a detailed description of the different imbalance metrics as well as the numerical weather prediction system and the experimental set-up. Section 4.4 describes the findings of the different metrics and compares the diagnostics. Finally, Section 4.5 summarises our findings and outlines further research that could be done.

4.3 Methods

4.3.1 Experimental Set-Up

Data Assimilation system

For this research, we use a near-operational set-up of the Km-scale ENsemble Data Assimilation (KENDA) system developed at DWD (Schraff et al., 2016). The KENDA system implements an LETKF data assimilation scheme (following Hunt et al., 2007) that couples with ICON-D2, a non-hydrostatic regional NWP model (Zängl et al., 2015). In this case, version 2.6.4 of ICON-D2 was used. The model is based on an icosahedral triangular grid, our experiments are calculated on the R19B07 grid which corresponds to a horizontal grid spacing of $\Delta x \approx 2$ km. The model domain is depicted in Figure 4.1 and covers Germany and neighbouring countries (grey box in Figure 4.1). The software "ICON-tool" provided by DWD was used to interpolate the data from the triangular grid to a regular latitude-longitude grid on rotated pole coordinates. To avoid problems with the margins of the model domain, the evaluation of the different experiments is performed in a sub-domain centered on Germany (blue box in Figure 4.1). The lateral boundary conditions are interpolated from the ICON-EU NWP model with $\Delta x \approx 7$ km. There are 65 terrain-following model layers in the vertical. The system uses a one-moment cloud microphyiscs scheme.

The experiments performed in this study are based on a reference experiment, which is close to the operational set-up as described by Schraff et al. (2016). The reference experiment simulates June and July of 2021. It comprises an ensemble of 40 members. LHN nudging is applied to each member every 20 seconds in the dynamic core.

LETKF data assimilation is performed every hour. Conventional observations, radar reflectivities and visible satellite images are assimilated (0.6 µm channel). The conventional observations are assimilated using adaptive horizontal localisation ranging from 50 to 100 km, chosen such that the number of assimilated conventional observations per analysis grid point stays constant (cf. Schraff et al., 2016). The radar data have a fixed localisation length scale of 16 km, and the satellite observations have a fixed localisation length scale of 35 km. Both parameters reflect the standard setting in the operational ICON-KENDA set-up and have been found to yield most beneficial impact in the forecast (Scheck et al., 2020). As in the operational set-up different techniques of covariance inflation are combined, namely multiplicative inflation, additive inflation and relaxation to prior perturbations (see Schraff et al., 2016). Hydrostatic balancing of increments is active (Potthast, 2019). Not all model variables are updated by the LETKF. In this study, pressure, temperature, humidity, horizontal wind (u, v), cloud liquid water and cloud ice water are updated, but the vertical wind is not updated at the analysis time. However, it is strongly constrained by the horizontal convergence of the wind. In the reference experiment the LEKTF increments are added using the incremental analysis update (IAU).

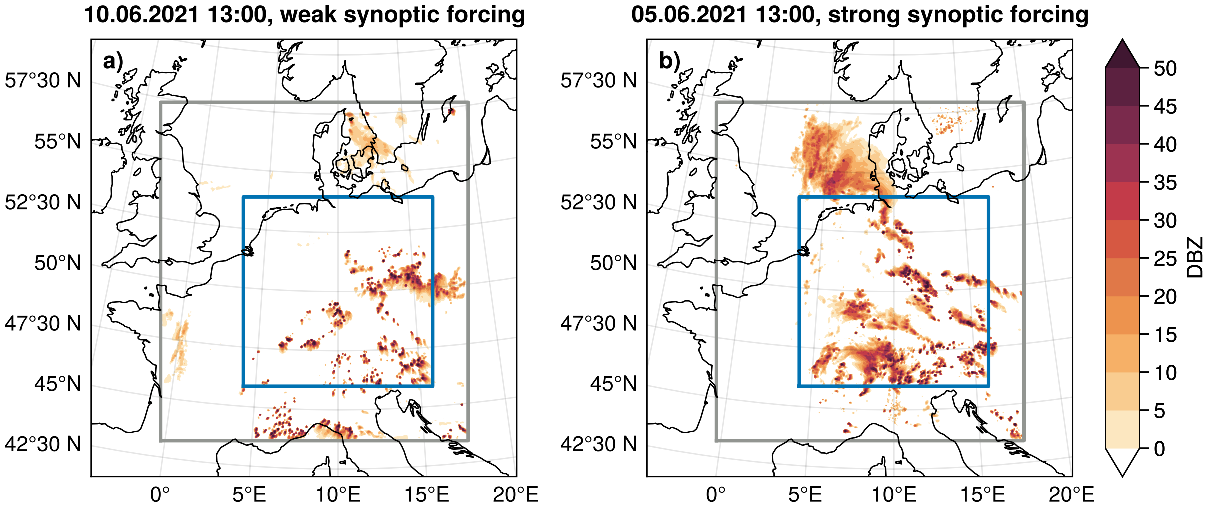


Figure 4.1: Model data from one member of the control run (noDA, member 1) of column maximum radar refelctivity for both cases. Panel a) shows Case 1, a weak synoptic forcing situation. Panel b) shows Case 2, a strong synoptic forcing situation. The grey box indicates the ICON-D2 domain (defined on rotated pole-coordinates). For the imbalance evaluation only the region in the blue box is considered.

Experiments

For our experiments we computed a set of analyses using different ICON-KENDA settings. For all the experiments, we use the initial conditions of the reference experiments as described in the previous section. From these initial conditions a background forecast is computed. In our case, we always start the background forecast at 12:00 UTC, and an analysis is computed at 13:00 UTC. The time frame that is of interest to us lies between 13:00 UTC and 13:50 UTC.
The experiments are designed as follow:

1. **The control experiment (noDA)**: The background forecast is initialised at 12:00 UTC from the reference experiment, the data assimilation step at 13:00 UTC is omitted, and the background simulation continues to run throughout the next hour. Latent heat nudging is switched off from 12:00 UTC onwards. This experiment is used as a basis to assess the effect of the LETKF and to characterise the natural level of imbalance in the system.

2. **The LHN experiment (LHN)**: This experiment explores the effect of LHN on the imbalance of the model state. The experiment is set up in the same way as the

noDA experiment, with the only difference that LHN runs until 13.00 UTC and is then switched off.

3. **The LETKF experiment (LETKF)**: The background forecast is initialised at 12:00 UTC from the reference experiment. LETKF data assimilation of conventional, radar, and visible satellite observation is performed and produces a new analysis at 13:00 UTC. At 13:00 UTC the first-guess forecast is initialised from the analysis. Latent heat nudging is switched off from 12:00 UTC onwards.

4. **The incremental analysis update experiment (IAU)**: The experiment is set up the same way like the LETKF experiment, with the only difference that the analysis increments computed for 13:00 UTC are not added all at once at 13:00 UTC but added incrementally over a time frame of 10 minutes, from 12:55 to 13:05. The IAU is applied uniformly through the 10 minutes. The time step of the updates corresponds to the dynamic time step of the model. This adheres to the near-operational set-up of ICON-KENDA as described in the previous section, except that LHN is switched off from 12:00 UTC onwards.

To save computational costs, we run each experiment using a 10-member ensemble.

Case studies

All experiments were performed on two different dates, with two distinct synoptic situations: Case 1 is a weak synoptic forcing case initialised at 10 June 2021 at 12 UTC and Case 2 is a strong synoptic forcing case initialised at 5 June 2021. The classification is taken from Puh et al. (2023). In their study the large scale forcing of three months in summer of 2021 was classified based on the convective adjustment time scale.

Figure 4.1 shows the radar reflectivity for both cases at 13:00 UTC. Panel a shows the weak synoptic forcing case, which is characterised by smaller convective features and high spatial variability in precipitation. The convective adjustment time scale in this case is longer, meaning that the large scale flow has minor influence on the removal of CAPE and convection is controlled by local triggering mechanisms. Panel b shows the strong synoptic forcing case, where the convective adjustment time scale is short, meaning that the large scale flow has a dominant influence on the local convection resulting in larger convective features and higher domain-averaged precipitation rates.

4.3.2 Measures of Imbalance

Surface Pressure Tendencies

In a variety of studies the domain averaged absolute surface pressure tendency (DPSDT) is used to investigate the degree of imbalance in the model introduced through data assimilation (see e.g. Houtekamer and Mitchell, 2005; Greybush et al., 2011; Lei et al., 2012; Ge et al., 2022). However, surface pressure tendencies offer only a limited view on the degree of imbalance in the atmosphere. Perturbations in surface pressure are determined by the mass convergence integrated over the whole atmospheric column above, thus relatively large values of convergence and divergence within the atmospheric column may cancel each other out (Lynch and Huang, 1992). So not all gravity wave motions will be reflected in the surface pressure tendency, while at the same time, other fast transient motions, such as acoustic wave may contribute. It is therefore an open question whether surface pressure tendencies are a good measure for the non-linear motions and non-hydrostatic (internal) gravity waves that have been identified with imbalance in previous studies.
The domain averaged surface pressure tendencies for one time step are computed as:

$$DPSDT = \frac{1}{N} \sum_{x_i, y_i} \left| \frac{dp_s}{dt} \right|, \tag{4.1}$$

where N is the number of model grid points (x_i, y_i) that are considered and $\frac{dp_s}{dt}$ is the time derivative of the surface pressure (p_s).

Vertical Motion Diagnostic

In this study, we will apply the vertical motion diagnostic (VMD) developed by Lange et al. (2017) to our data assimilation experiments. The idea behind this diagnostic is to analyse vertical velocities on a horizontal level in order to detect signals that indicate the presence of gravity wave noise. The diagnostic should be applied to a level where gravity wave noise that could be associated with the triggering of spurious convection is likely. A simple method would be to compute the spatial variance of the vertical velocity field, however, this could be misleading because the variance in a convective weather situation will be dominated by high vertical velocities in the convective updrafts. The regions that are most sensitive to gravity wave noise, however, are not the convective updrafts but the near-convective environment. As an illustration, one can keep in mind that a typical updraft in a convective cell is on the order of 10 ms^{-1}, while typical amplitudes of gravity waves are on the order of 1 ms^{-1}. For this reason Lange et al. (2017) designed a masking algorithm that distinguishes between the dominant convective updrafts, the variance in the near-convective environment and more distant from the convection. The partitioning of the variance is achieved by utilising radar reflectivity (DBZ) threshold values, which serve as a basis for defining three distinct masking regions:

- The **inner mask** is defined as the region inside the convection. It has a three-dimensional structure and is confined by a threshold reflectivity of 5 DBZ.

- The **vicinity mask** is defined as the region in the near-convective environment. It is based on a threshold of 5 DBZ in the two-dimensional column maximum radar reflectivity field. This boundary is then dilated in the horizontal by a distance of 16 km which ensures that the typical size of a cold pool in the mid-latitudes is also enclosed in the vicinity region. The grid points of the inner mask are excluded from the vicinity.

- The **outer mask** is defined as all grid points that are not inside the inner or inside the vicinity mask.

These definitions and threshold values exactly follow Lange et al. (2017). However, there is an important difference to Lange et al. (2017) when it comes to the application of the algorithm. In this study, we will apply the diagnostic to experiments based on a near-operational NWP system, whereas Lange et al. (2017) applied their algorithm to idealised simulations, which allowed them to compare their data assimilation runs to a nature run. Therefore, they defined the three different masking regions on the basis of the nature run. In our case, the masking regions will be computed individually for each experiment, since the different data assimilation methods will change the convective cloud fields relative to the noDA-reference run in different ways, requiring different masking regions.

We will apply the VMD algorithm to a horizontal layer at 5 km height, which is far enough away from the turbulent boundary layer or orography, but still on a level where spurious convection could be triggered. For the sake of comparability between the VMD and the WTG diagnostic, we chose geometric height as the z-coordinate.

Deviations from the Weak Temperature Gradient Approximation

As a third measure of imbalance, we will advance the idea of a new diagnostic based on the assumption that convective motions in the mid-latitudes approximately satisfy the weak temperature gradient (WTG) approximation, which means that the vertical upward motion is determined by the requirement that the resulting adiabatic cooling balances exactly the latent heat source. This was predicted by Klein (2010) and later verified by Craig and Selz (2018), who also proposed this as a potential balance principle for the mesoscale.

The vertical velocity satisfying the WTG approximation (w_{wtg}) is computed as:

$$w_{wtg} = \frac{q_\theta}{\partial_z \theta_0}, \tag{4.2}$$

where $q_\theta = \frac{dq_\theta}{dt}$, is the diabatic heating rate and $\partial_z \theta_0$ is the vertical gradient of a background potential temperature field (θ_0). For the derivation of Eq. 4.2 we refer the reader to Craig and Selz (2018). We will call the difference between the actual vertical wind component (w) and the wind predicted by the WTG approximation (w_{wtg}) the residual (w_{res}), defined as:

$$w_{res} = w - w_{wtg}. \tag{4.3}$$

w_{res} then indicates the portion of the vertical wind that is not balanced by heat sources. To compute w_{wtg} all relevant variables were interpolated to constant geometric height levels. In our model set-up the variable tt_lheat was used as the heating rate q_θ, which collects all heating rates that are present in the model, including the microphysics parameterisation.

In the control experiment it is observable that the departure from the WTG balance (w_{res}) is correlated with the amount of precipitation in the domain. It is therefore desirable to normalise w_{res} by the precipitation rate. Figure 4.2 shows the relation between the domain averaged absolute mean of w_{res} at one horizontal layer at 5000 m height and the domain averaged precipitation rate, for the time series from 13:01 to 13.50 UTC of the noDA runs. The individual colors indicate the individual ensemble members. The blue line indicates a linear fit. To separate the amount of imbalance due to enhanced precipitation from the imbalance likely due to excess gravity wave noise, we apply a

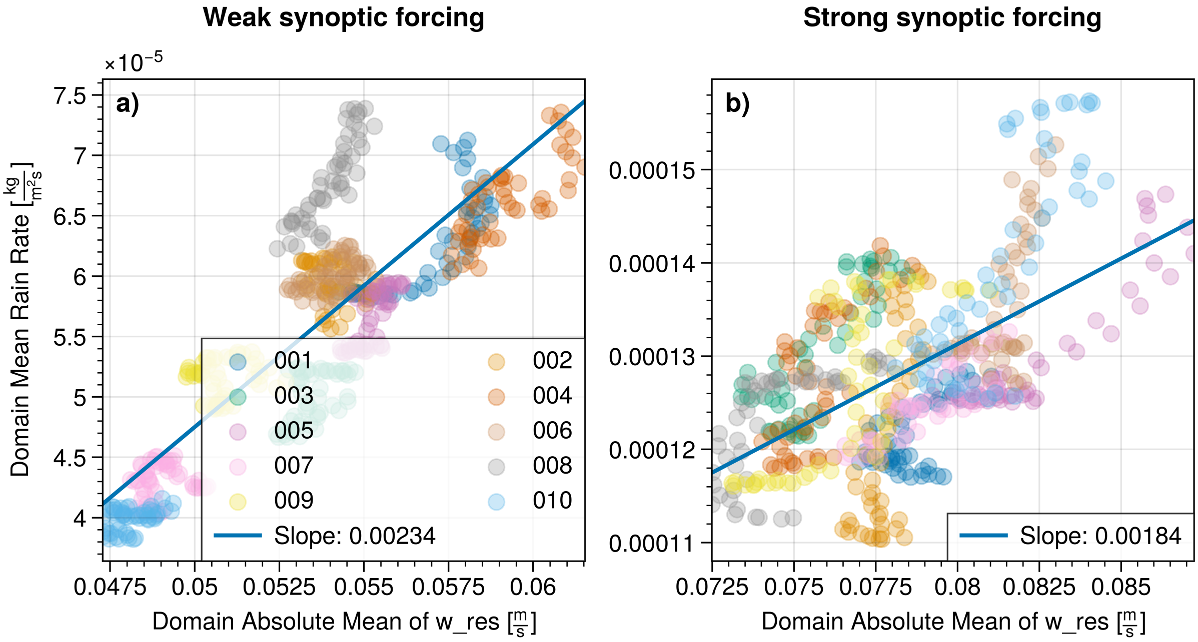


Figure 4.2: The domain absolute mean of the WTG residual (w_{res}, x-axis) over one horizontal model level at 5000 m versus the domain mean precipitation (y-axis) for the control run (noDA). The color of the dots indicates the number of the ensemble member. The individual dots of one member represent one time step between 13.01 and 13.50. The blue line indicates the linear fit. The slope of the fitted line is used for the correction of w_{res}. Panel a) shows Case 1, the weak forcing case and Panel b) shows Case 2, the strong forcing case.

linear correction of the form,

$$w'_{res} = w_{res} - \frac{1}{\gamma}(precip_{exp} - precip_{noDA}), \tag{4.4}$$

where γ indicates the slope of the fitted line (blue line in Figure 4.2) and $precip$ stands for the precipitation rate in the DA experiments, with $exp \in$ [LETKF, IAU, LHN]. Note that a similar correction is not required for the VMD diagnostic, since the area of the inner and vicinity regions expand as the area occupied by precipitating convection increases.

4.4 Results

In this section we apply the three different imbalance measures to the different data assimilation experiments for the weak and strong forcing cases. For the evaluation we focus on the time frame from 13:01 UTC to 13:50 UTC for both cases. The period starts at 13:01 because quantities such as the surface pressure tendency cannot be computed at the analysis time in the DA experiments. The results are always shown for the first ensemble member of the simulations, since they are similar for all members.

A first impression of the impact of the different data assimilation methods can be obtained from Figure 4.3, which shows the time evolution of the spatially averaged precipitation (sum of convective and microphysical precipitation rates) for the different experiments (different colored lines) and for both cases (Panel a) and b)). Generally, it is observable that there is more rain in the strong forcing case (Case 2) than in the weak forcing case (Case 1). The highest amount of precipitation is found in the LHN experiments, and the difference persists through the first 50 minutes of the new forecast cycle. In general, the observed precipitation (not shown) lies in between the predictions of the LHN and LETKF experiments. This suggests that LHN is more effective in introducing precipitation to the model, and that there is probably a dry bias in the model itself.

For the LETKF experiment we see enhanced rain rates with respect to the control run (noDA) in the first ten to fifteen minutes after the analysis time, then decreasing. At longer lead times the precipitation amount is less than the noDA experiment for the weak forcing case, and greater for the strong forcing case. After the updates have finished at 13:05, the precipitation in the IAU experiment is very similar to that of the LETKF method.

4.4.1 Surface Pressure Tendencies

Figure 4.4 shows the domain-averaged absolute mean of surface pressure tendencies of the first ensemble member. Panel a) shows the result for Case 1 and Panel b) shows the result for Case 2. The signatures of the surface pressure tendencies are similar in both cases. The DPSDT level associated with balanced dynamics can be identified from the control run, which remains constant over time at around 2 Pascal per second. The sawtooth pattern in the curve is likely produced by the radiation scheme in the simulation which is called at two minute intervals in the model simulation.

The LETKF runs show large initial surface pressure tendency, which decays exponentially to the level of the control run. One minute after the analysis time the surface pressure tendency lies at 22 Pascal per second in Case 1 (the weak synoptic forcing case) and 34 Pascal per second in Case 2 (the strong synoptic forcing case), in both cases at least an order of magnitude larger than the noDA level. The pressure tendency level decays

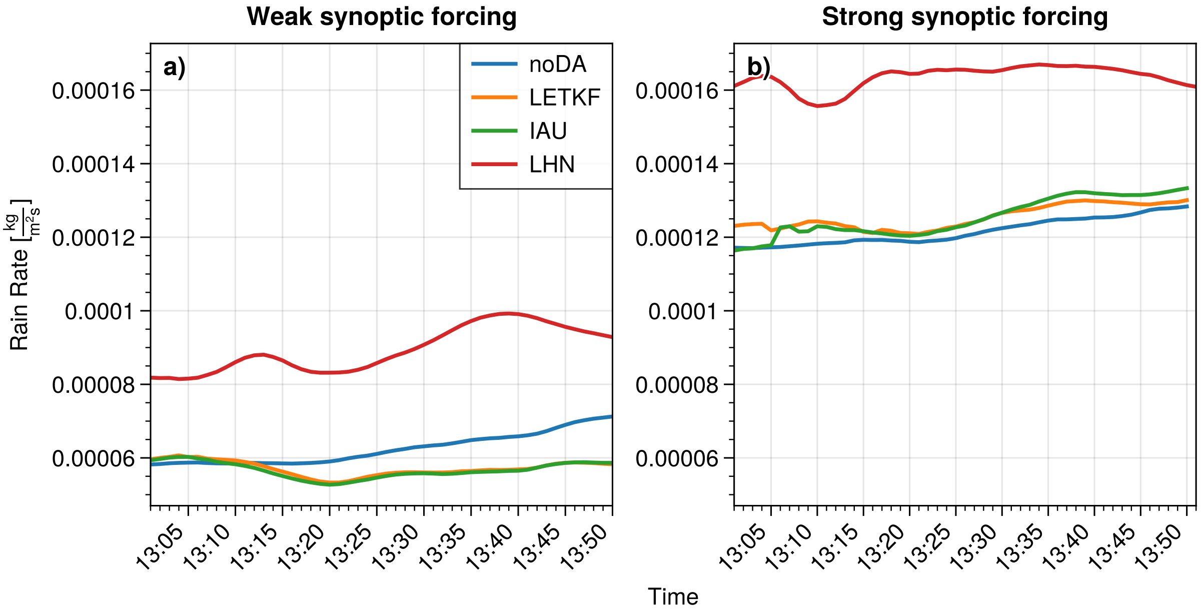


Figure 4.3: Total precipitation rate (convective and microphysical) as a function of time for the first ensemble member. The different colored lines indicate the different experiments.

roughly exponentially with a relaxation timescale of 5 to 10 min. The IAU experiments show significantly smaller surface pressure tendencies. After the first 5 minutes, during which the increments are still being added, the level is fairly constant for an additional ten minutes, then decays with a similar exponential timescale to the LETKF experiments. By this measure the LETKF introduces a very noticeable degree of imbalance, and the IAU method is effective in reducing it substantially. Surprisingly, the LHN experiment does not seem to create significant imbalance. The small increase in surface pressure tendency relative to the noDA experiment that persists throughout the period is likely associated with increased precipitation amounts (see Fig. 4.3), rather than a less balanced model state.

4.4.2 Vertical Motion Diagnostic

As described in the previous subsection, a crucial point when analysing imbalance and noise introduced through data assimilation is to determine whether or not the detected noise is dynamically relevant and how it will affect the spin-up of the model and the potential evolution of spurious convective cells. Lange et al. (2017) found that the surface

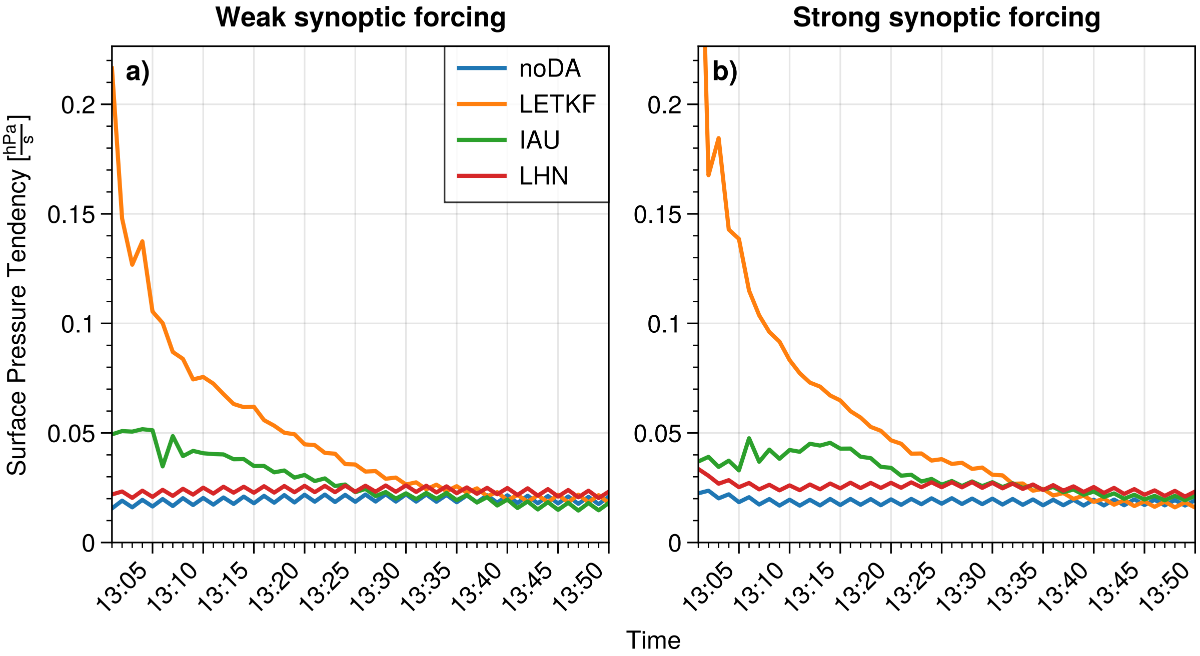


Figure 4.4: Mean absolute of surface pressure tendencies as a function of time for the first ensemble member. Panel a) shows Case 1, the weak synoptic forcing case and Panel b) shows Case 2, the strong synoptic forcing case. The different colored lines indicate the different experiments.

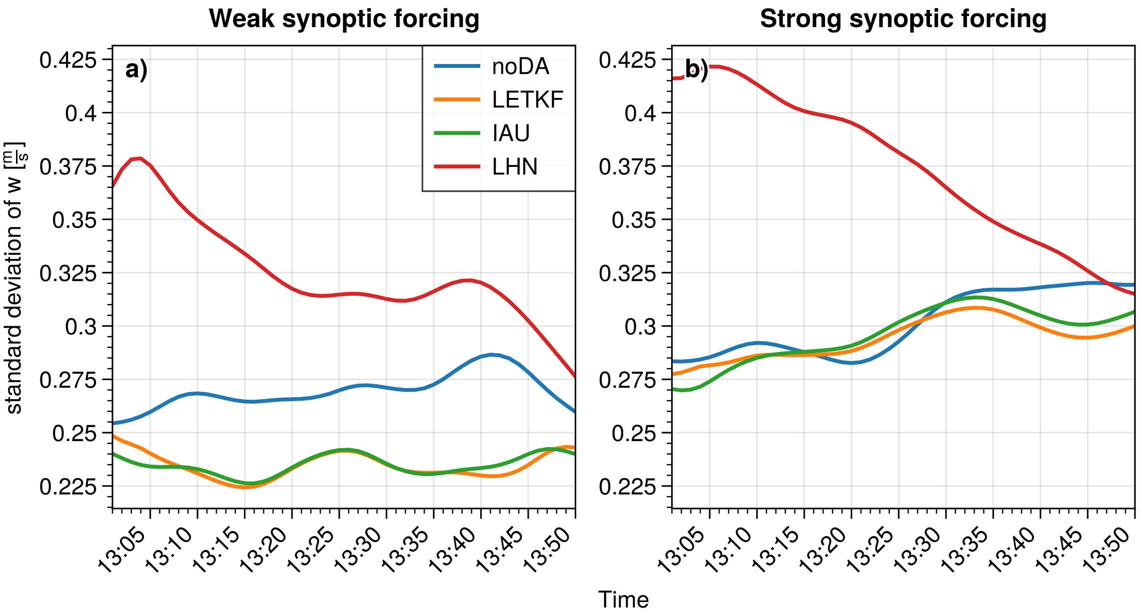


Figure 4.5: Standard deviation of the vertical velocity field at 5 km height as a function of time for the first ensemble member. Panel a) shows Case 1, the weak forcing case and Panel b) shows Case 2, the strong forcing case. The different colored lines indicate the different experiments.

pressure tendencies were not a reliable predictor of the formation of spurious convective cells in their idealised sensitivity experiments, and proposed a measure that was more directly related to vertical motions that could trigger convection. As noted previously we will apply the vertical motion diagnostic of Lange et al. (2017) for the horizontal model layer at 5 km. Similar results are seen at all levels between ~ 3 to 6 km height. Below 3 km, the results are complicated by the proximity of orography and turbulence in the boundary layer.

Before considering the vertical velocity variance in the different mask regions, we show in Fig. 4.5 the standard deviation of the vertical velocities averaged over the entire evaluation domain. The LETKF and IAU experiments show less variance than the reference, while the LHN experiment show substantially more. However, this is not primarily an indicator of different degrees of imbalance. Rather, the differences between the experiments is correlated with the precipitation amounts (Fig. 4.3), except for the decrease of vertical velocity variance over time seen for the LHN experiment.

Figure 4.6 shows the standard deviation of vertical velocity in the three mask regions. The hypothesis that the total vertical velocity variance is mainly influenced by the amount of convection is supported by the strong resemblance between the overall vertical velocity variance (Fig. 4.5), and the variance confined to the inner region of the convection (first row of Fig. 4.6). There is a strong tendency for LHN to increase the intensity of the convective updrafts, while the LETKF seems to weaken the updrafts somewhat in the weak forcing case. However, much of this is a balanced response to changes in the heating rates, rather than an indication of transient gravity waves that would be associated with imbalance. In contrast, Lange et al. (2017) argued that increased vertical velocity variance in the vicinity of convection was a more appropriate measure of the gravity wave variability that might be spuriously increased by data assimilation. This is shown in the second row of Fig. 4.6. The LETKF shows increased variance in the vicinity region in the first 5 - 10 minutes after the assimilation increments are added, indicating increased imbalance. The amplitude of this increase is strongly reduced and shifted by the IAU. The LHN experiment produces a large initial increase in variance in the vicinity region, which decays throughout the period. In fact, all of the DA experiments show a decrease with time of vertical velocity variance in the vicinity region, which is consistent with an adjustment towards balance. Note that the levels do not return exactly to that of the reference experiment at the end of analysis period. It is possible that the change in the area of the vicinity mask is not precisely compensating the change in the overall amount of convection resulting from the data assimilation. Finally, the contribution of the vertical velocity variance in the outer region (last row of Fig. 4.6) is small in comparison to the other two regions. Overall, and in contrast to the surface pressure tendency diagnostic, the VMD measure suggests that LHN produces the largest imbalance, while the LETKF produces a significant but smaller imbalance. The IAU appears effective in reducing the production of unbalanced motions.

4.4.3 Deviations from the Weak Temperature Gradient Approximation

Figure 4.7 shows the domain-averaged absolute mean of the weak temperature gradient vertical velocity residual corrected by the amount of precipitation in the individual experiments (w'_{res}), at 5 km height (see Section 4.3). A low residual indicates that the vertical velocity in the domain is predominantly a slow response to diabatic heat sources rather than transient gravity waves, and is, in this respect, more in balance. As in the DPSDT diagnostic, the control run has the lowest residual, meaning that it is the most balanced state; moreover, it stays approximately constant throughout the forecast. In the strong

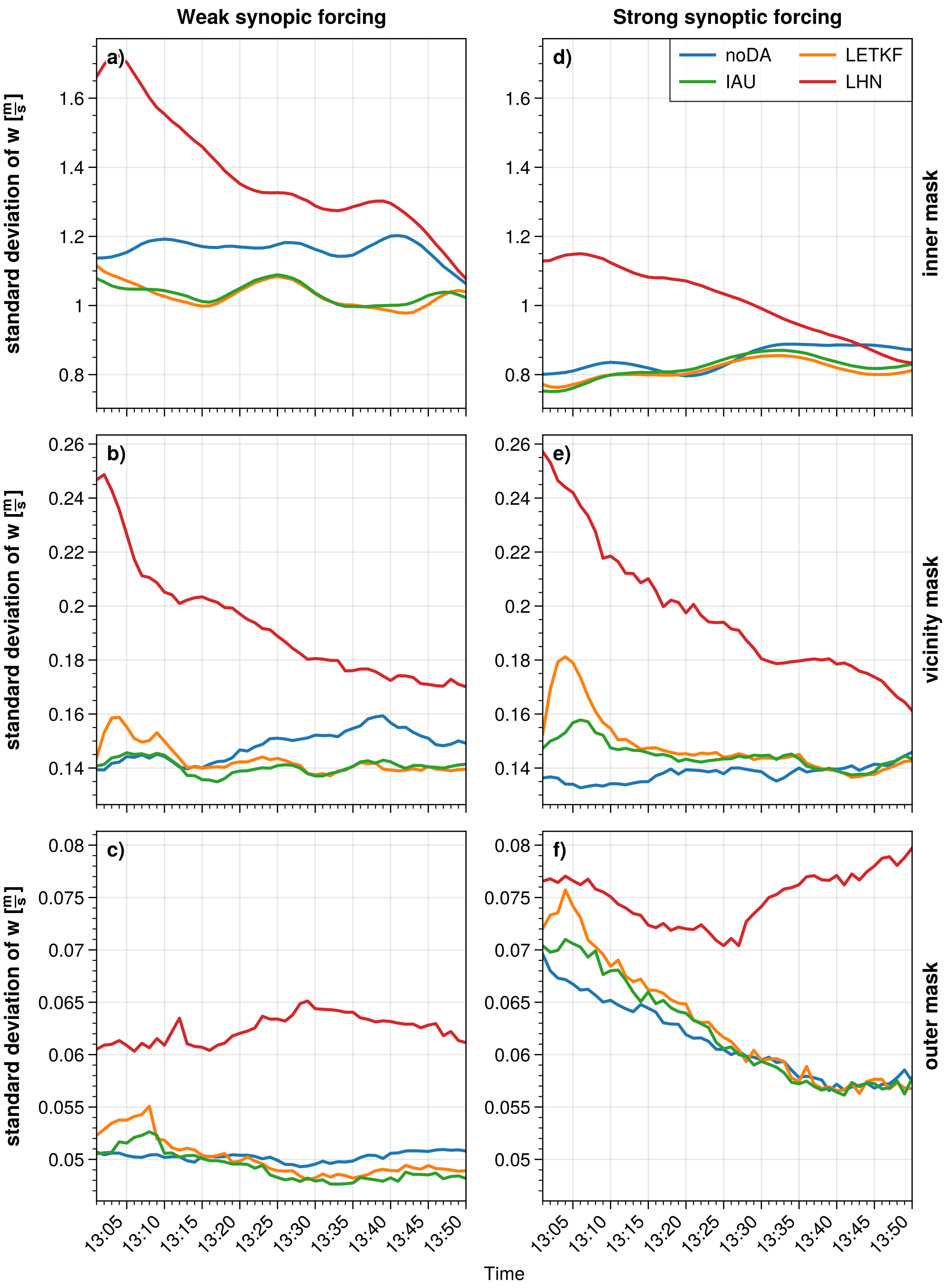


Figure 4.6: Standard deviation of the vertical velocity field at 5 km height as a function of time for the first ensemble member in the different masking regions (inside the convective updraft, in the vicinity and outside the convection). Panel a),b) and c) show the results for Case 1, the weak forcing case and Panel d),e) and f) show the results for Case 2, the strong forcing case. The different colored lines indicate the different experiments.

forcing case, both LETKF and IAU show elevated levels of w'_{res}, lasting to 30 minutes. The IAU experiment shows a reduced level of imbalance compared to the LETKF, but the difference is not as dramatic as seen in the previous imbalance measures. In the weak forcing case, only the LETKF shows slightly elevated w'_{res} in the first 5 minutes and for IAU it even starts slightly below the noDA run. The LHN experiment shows largest w'_{res}, which decreases over a period of 40 minutes or more. This is unsurprising, since as noted in the introduction, heating increments are introduced without consistent changes to other model variables, leading to large imbalanced motions as the other variables adjust towards a balanced state.

All of the experiments relax towards the reference level over time, showing that the normalisation to the total precipitation amount is successful in removing the effects of the changes in amount of convection from the different DA schemes. There appears to be a slow oscillation in the amplitude in the weak forcing case, which can be related to variations in the total precipitation in that experiment (Fig. 4.3), rather than changes in the vertical velocities themselves. The conclusions regarding the amount of imbalance introduced by the different DA methods are qualitatively similar to those of the VMD diagnostic, and different from the results of the DPSDT measure.

4.4.4 Relations between the Different Measures

The three measures show clear differences in the relative degree of imbalance produced by the different DA schemes. The surface pressure tendency shows the largest imbalance resulting from the LETKF, and almost no imbalance from LHN. Meanwhile, the other two measures, based on vertical velocity in the troposphere, show the greatest imbalance from LHN, with a smaller but still significant impact in the LETKF experiments. The vertical motion diagnostic and the WTG departures show similar results, which suggests that they are capturing the same physical phenomenon. On the other hand, the completely different results from surface pressure tendencies suggest that this diagnostic is responding to something different.

Further insight can be obtained by examining the spatial patterns of the imbalance diagnostics. We describe only Case 2, since the analysis of Case 1 yields similar conclusions. The upper row of Fig. 4.8 shows results for the reference run while the lower row presents the LETKF experiments as an example of the DA results. All snapshots are taken at 13:02 UTC, i.e. two minutes after the analysis time. The first column depicts

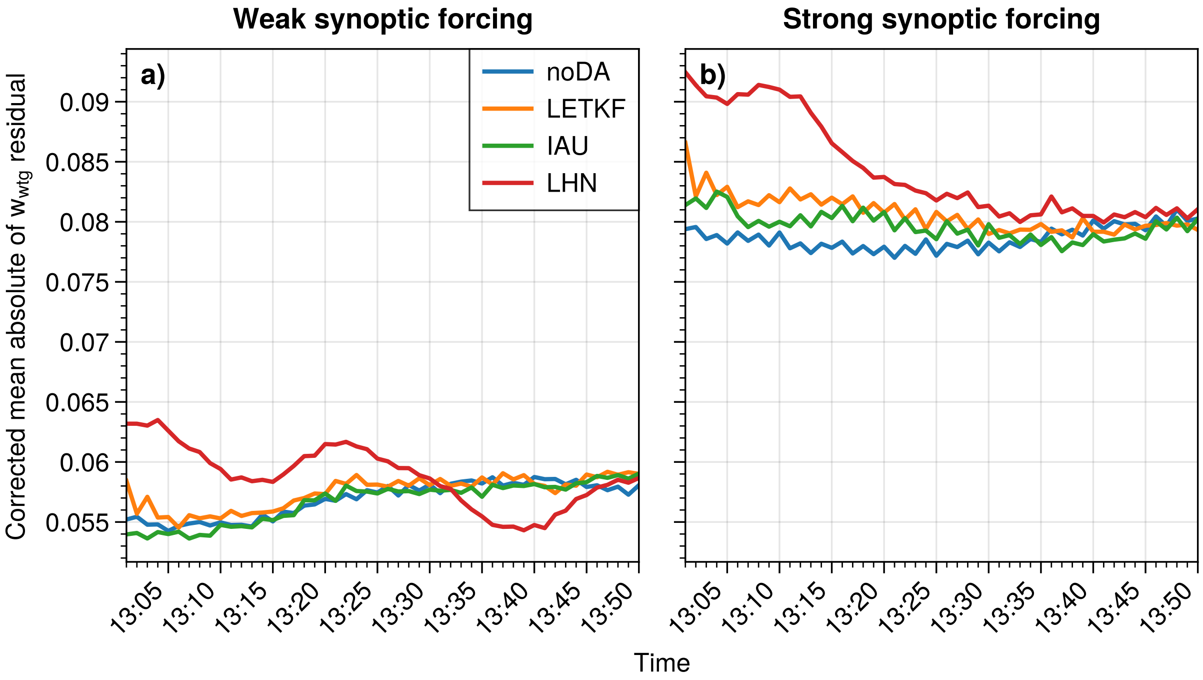


Figure 4.7: Mean absolute of the corrected WTG residual ($w_{res} = w - w_{wtg}$, corrected by the amount of precipitation as compared to the control run) at 5 km height as a function of time for the first ensemble member. Panel a) shows Case 1, the weak forcing case and Panel b) shows Case 2, the strong forcing case. The different colored lines indicate the different experiments.

the surface pressure tendencies, the center column depicts the vertical velocity in the vicinity mask at 5 km height and the third column depicts the WTG departure w_{res} (no precipitation correction).

In the reference run (first row of Fig. 4.8), all three measures have their largest values in areas of active convection. The surface pressure tendency has largest amplitudes at the locations of the convective clouds, but sometimes the tendencies are positive and sometimes negative. This is presumably related to the convective lifecycle, with falling surface pressure below growing updrafts turning to increasing surface pressure as convective downdrafts develop. The amplitude of the perturbations decreases away from the convective cores. Considering now Panels b and c of Figure 4.8, the highest values of tropospheric vertical velocity are located in the convective cores, but these are masked out in the VMD diagnostic. The vicinity region is dominated by gravity wave motions that in some cases form rings around the convective cores. Outside of the convective cores, there is little diabatic heating and the WTG departures are similar to the total vertical motions. Inside the convective cores, the WTG approximation overestimates the updraft speed (Craig and Selz, 2018), so the WTG departure is negative with reduced amplitude compared to original vertical field. At least in the reference experiment, all three imbalance measures are consistent with each other in showing that the degree of imbalance is closely coupled to the strength of the convective motions.

In the LETKF experiment (second row of Fig. 4.8), on the other hand, the imbalance measures show different spatial patterns. In particular the surface pressure tendency shows large-scale perturbations that extend across the domain. In contrast, the VMD and WTG diagnostics show similar spatial patterns to those found in the reference experiment but with more intensity. In the later two diagnostics, the measured imbalance is directly related to the convection.

The structures present in the surface pressure tendency after the analysis (Fig. 4.8d) are dominated by scales from about 20 km to more than 50 km, which is considerably larger than the small convective cells visible in the other panels. While the amplitudes of the structures tend to be larger where the density of convective cells is higher, variability on these scales is also found in regions without convective activity (e.g. upper left or upper right corners of Fig. 4.8d). The surface pressure tendency distribution thus cannot fully be explained by direct influence from the convective cells propagating into the rest of the domain in the short time interval of two minutes considered here. An additional mechanism not directly relying on convection must be active.

A possible explanation is related to the impact of localisation on the analysis pressure field. The horizontal localisation scales depend on the observation type (see Section 4.3) and are typically between 16 km and 50 km. The LETKF analysis is formulated as a linear combination of first guess ensemble members, with weights varying on the localisation scale. For relatively smooth fields like pressure, which in the absence of convection varies on length scales larger than the horizontal localisation scales adopted here, the analysis will therefore be dominated by structures on these localisation scales. In the forecast, pressure adjustment processes will cause surface pressure tendencies on the same scales. Even without convective activity one would therefore expect pressure tendencies to be created on scales like those visible in Figure 4.8d as a result of localisation. The surface pressure tendency metric will measure both the imbalance caused by localisation in the pressure field and pressure perturbations created at the convective cells.
These results suggest that the different diagnostics contain complementary information. As noted previously, Lange et al. (2017) has suggested that the vertical motion diagnostic is a better predictor of the triggering of spurious convective cells than surface pressure tendencies, and therefore should be considered in convective-scale DA. Similar results are also obtained using WTG departures, and this latter method has some advantages. First, the weak temperature gradient approximation can be theoretically justified as a suitable balance for the convective scale (Craig and Selz, 2018; Klein, 2010; Hirt et al., 2023). Secondly, the WTG departure is computationally more efficient since it uses single level data, and it does not require the definition of different masks.

4.5 Conclusions

In this study, we evaluated three distinct measures of imbalance at the convective scale, each grounded in different physical rationales. Our primary objective was to discern differences among these measures and explore how they respond when different data assimilation methods are applied. The three different imbalance measures that were tested are: (1) surface pressure tendencies, (2) vertical velocity variance in the vicinity of convective clouds and (3) departures from the vertical velocity that is prescribed by weak temperature gradient balance. The strategy was to apply them to forecasts produced with three data assimilation schemes: LHN, the LETKF and the LETKF in combination with IAU. We chose these three DA algorithms because, by design, different degrees of imbalance in the respective analyses can be expected.

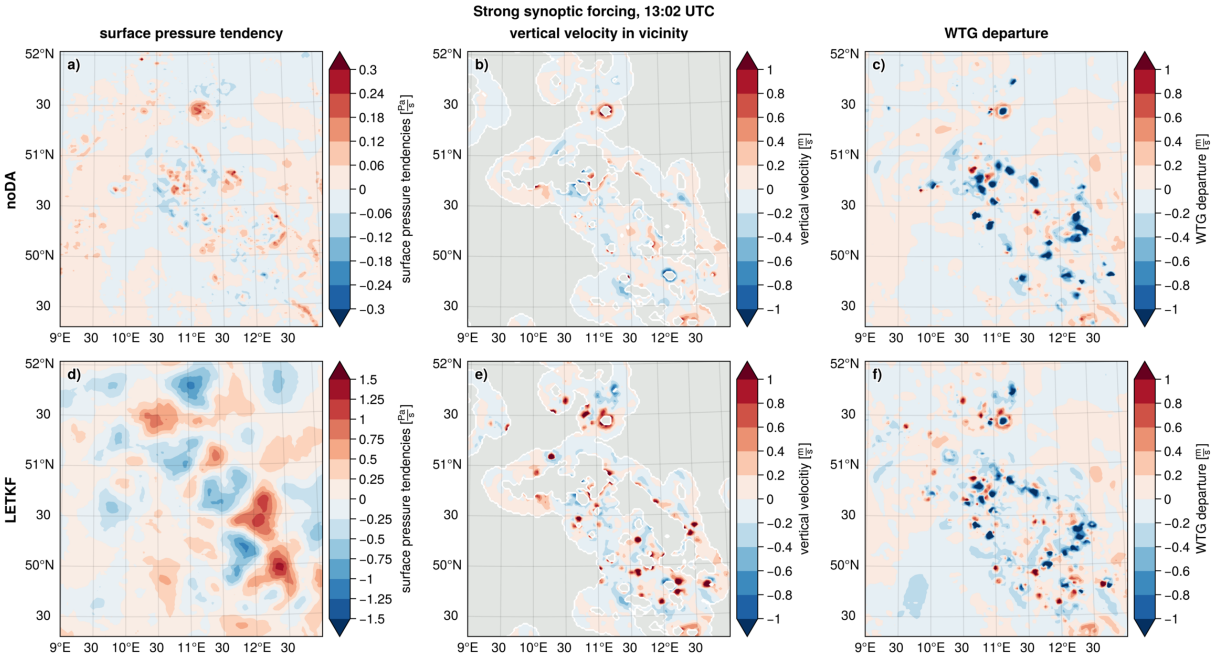


Figure 4.8: Spatial maps of the different imbalance measures in a small subdomain. The upper row the control run (noDA), the lower row shows the LETKF run. The snapshots are taken at 13:02, two minutes after the re-initialisation in the LETKF run. Only the strong synoptic forcing case (Case 2: 05.06.2021) is shown. Panel a) and d) show surface pressure tendency. Panel b) and e) show the vertical velocity inside the vicinity mask at 5 km height. The grey shading indicates regions outside the vicinity mask. Panel c) and f) show the departure from WTG residual (w_{res}) at 5 km height at 13:02.

The first key result shown in this work is that surface pressure tendencies seem to measure a different kind of imbalance as compared to the two other measures. In terms of DPSDT-imbalance the LETKF experiments exhibit the largest amount of imbalance, while the other two measures indicate largest imbalance in the LHN experiments, followed by the LETKF and the IAU runs. Moreover, we could observe that the signatures in surface pressure tendencies show much larger spatial patterns and faster decay time scales as compared to the other two measures. This appears consistent with the hypothesis that non-hydrostatic gravity wave noise is not detected by the DPSDT and the external gravity wave mode dominates the signal in the surface pressure.

The second main outcome of this study is that departures from WTG balance seem to be a useful measure for imbalance on the convective scale. The results are very similar to those obtained with the vertical velocity variance diagnostic. For computational efficiency we recommend the use of WTG departures over that of the partitioned vertical velocity variance. Moreover, the vertical velocity variance has several thresholds and design choices that might influence the results, which are not required for the WTG diagnostic. A potential disadvantage of the WTG diagnostic is that it requires knowledge of the total heating rate in the model, which is not a standard output variable in ICON-KENDA at the time of writing . For both diagnostics we employed model variable interpolation to constant geometric height, aligning with the theoretical derivation of the weak temperature gradient (WTG) approximation in Craig and Selz (2018). This does not appear to be necessary, as we also tested applying the diagnostics directly on model levels, yielding comparable results and avoiding the computationally expensive interpolation to constant height (results not shown).

When it comes to the behavior of the data assimilation algorithms, we find that the WTG and VMD measures are more in line with our physically-based expectation about the imbalance introduced by the different algorithms. In particular, these diagnostic show that the imbalance produced by LHN was significantly larger than for the LETKF methods. The incremental update method (IAU) decreases the imbalance introduced by the LETKF efficiently in our experiments. In terms of flow-dependence we did not see significant differences between the two different synoptic-forcing cases. In summary, this suggests that the WTG departures offer important complementary information about convective-scale imbalance introduced through data assimilation and should be used in combination with $DPSDT$ when investigating the behavior of a DA system.

Some significant limitations of this work should be noted however. First, the number of cases examined is very limited, and focused on forecasts of convection where imbalance is a major concern. Secondly, both the vicinity vertical motion and the recommended WTG departure diagnostic require a reference experiment to set the background level of imbalance. It might be sufficient in operational use, to obtain this from the background forecast at the analysis time, rather than integrating the reference forecast for an hour as was done here. Another important question is if and how the imbalances measured here impacts the subsequent forecasts, for example by triggering spurious convective cells. Unfortunately, while spurious convection is easy to identify visually in idealised simulations (Aksoy et al., 2009; Lange et al., 2017), this is not the case in the more realistic simulations here, where many factors lead to differences in the convective cloud field. Further work will be required to isolate the effects of imbalance as the forecast evolves nonlinearly. Finally, this study has only addressed questions of how the different measures quantify imbalance resulting from different DA methods. It remains to be determined if and how the imbalance impacts forecast skill for user-relevant variables such as precipitation.

In this study, we have tried to address the need identified in recent reviews (Gustafsson et al., 2018; Hu et al., 2022) for specialised techniques tailored to the convective scale to investigate the imbalance introduced by data assimilation. The diagnostics examined here can contribute to a better understanding of convective scale imbalance and pave the way for further research. For example, it may be interesting to use observation impact diagnostics, such as partial analysis increments (Diefenbach et al., 2023) or EFSOI (Kalnay et al., 2012; Hotta et al., 2017b) to investigate contributions to imbalance from different observation types and other components of the data assimilation system. In addition, the investigation of the interplay between imbalance and forecast skill is a compelling avenue for future research.

Chapter 5

Conclusions

In this thesis we have addressed two scientific challenges. The first one was how to compute observation influence in a computationally efficient way, such that the algorithm is suitable for monitoring the influence of a multitude of observations assimilated in a convective-scale data assimilation system. The second challenge was concerned with the topic of imbalance introduced through data assimilation and how to measure it. In the frame of this thesis two novel diagnostic tools, namely partial analysis increments and the weak temperature gradient imbalance metric, were implemented and assessed. In the following we will summarise the findings, respond to the research questions posed in the introduction of this thesis and discuss the main conclusions.

5.1 Summary and Main Conclusions

Research Question 1 *How can we evaluate the impact of various observations in a way that yields easily interpretable results and ensures computational efficiency?*

For convective-scale data assimilation (DA) there is potentially a vast amount of observational data available, such as from satellites, ground-based remote sensing instruments or human and economic activities. However, the assimilation of such complex observational data is challenging due to complex observation operators or simply the sheer amount of assimilated observations. Therefore, literature such as the reviews by Gustafsson et al. (2018) or Hu et al. (2023) suggests that there is need for novel observation influence and impact diagnostics in order to be able to monitor and evaluate observation networks and DA systems. Moreover, there is little knowledge about optimal assimilation settings, such as localisation length scales, cycling frequencies or spatial distributions, therefore sensitivity studies are required.

In this thesis we use the a regional modeling system of the German Weather Service, COSMO-KENDA, which consists of a nonhydrostatic convective resolving forecasting model that runs at a horizontal grid-resolution of $\approx$ 2 km, and therefore resolves deep convection explicitly, and a DA system that is based on the Local Ensemble Transform Kalman Filter (LETKF). The LETKF allows for explicit calculation of the Kalman Gain matrix and by this the contribution of every observation to the analysis field (partial analysis increment, PAI). Despite providing very valuable information, PAIs have been rarely used as diagnostics up to now. In Chaper 3, we demonstrate how to compute PAIs, providing a mathematical derivation from the LETKF equations. Further, we discuss potential applications and propose their use for optimising LETKF DA systems, in particular with respect to satellite DA and vertical localisation. For computational efficiency, we calculate the Kalman Gain based on ensemble analysis perturbations which are available in the standard LETKF output, and not the higher dimensional weight matrix, which would require considerable additional effort to write out. An approximation has to be made to take localisation into account. However, we demonstrate that this only introduces minor errors as the localisation function changes very gradually among nearby grid points. Furthermore, the influence of observations always depends on the presence of other observations and specific DA parameters. These parameters include the observation error, which determines the weighting of the observations and the localisation length scales, which define the horizontal and vertical extent of their influence. Yet, we show that the assimilation of other observations primarily decreases the magnitude of the influence, but does not change the overall structure of partial analysis increments significantly. This means that the calculation of partial analysis increments can be used as a computationally efficient diagnostic to investigate the three-dimensional influence of observations in the assimilation system on various analysis variables. We demonstrate how to leverage this diagnostic to detect if the influence of additional experimental observations (e.g. cloud-affected satellite observations) is in accordance with other observations (e.g. radiosondes) without conducting computationally expensive single-observations or data denial experiments. This allows for the detection of observations or observation groups that lead to opposing analysis increments which may be related to imperfect assimilation settings or systematic errors in the observations. Last but not least, we show that the calculation can be used to approximate the influence an observation would have when applying different localisation settings for optimising the assimilation system.

Research Question 2 *Which measures of imbalance are suitable for the convective scale?* The analysis update produced by DA is not necessarily dynamically consistent with the forecasting model. The most significant form in which this inconsistency manifests itself is as imbalance, which means that the prevailing balances of forces in the atmosphere are disturbed in the analysis. After initialisation from an imbalanced analysis the model will strive towards a balanced model state which may yield undesired adjustment processes in the form of transient gravity waves early in the forecast. For convective-scale DA systems it has been shown that a consequence of imbalance are also spurious convective cells, which may be triggered by gravity wave noise and have the potential to degrade the forecast. Therefore, it is important to develop methods to detect and correct for imbalance introduced through DA. While there are effective procedures to reduce imbalance on the synoptic scale, the situation on the convective scale is less clear. Nonlinearity as well as non-hydrostaticity play an important role in the formation of convection and there is no clear separation between fast and slowly evolving time scales. Therefore, the development of methods to counteract imbalances in convective-scale DA is subject to current research and even the question of how to measure imbalance on the convective scale is unresolved. In this study we investigate three different measures of imbalance that are potentially relevant for convective-scale DA: 1. surface pressure tendencies, 2. vertical velocity variance in the vicinity of convective clouds, 3. departures from the vertical velocity that is prescribed by the weak temperature gradient (WTG) approximation. To test the effectiveness of these measures, we apply them to three different DA schemes, namely latent heat nudging, LETKF and LETKF in combination with the incremental analysis update. These DA schemes are chosen such that different degrees of imbalance in the respective analyses are expected. The experiments are based on the ICON-KENDA numerical weather prediction system of the German Weather Service, in a near-operational set-up. Moreover, our experiments involve two different cases with two distinct synoptic situations (weak and strong forcing).

While all three imbalance measures detect differences in the behaviour of the different DA schemes, it seems that surface pressure tendencies diagnose a different type of imbalance than the vertical velocity variance and the WTG departures, which show very similar results. Surface pressure tendencies indicate that the LETKF update produces the strongest imbalances, whereas the other two measures suggest that LHN generates the most imbalanced analyses. This indicates that vertical velocity variance and WTG departures provide complementary information to the more conventionally used surface pressure tendencies in terms of convective-scale imbalance. For reasons of numerical efficiency, we propose the WTG departures over the partitioned vertical velocity variance. These findings pave the way for further investigations, such as examining the relationship between imbalance in different metrics and forecasting skill.

5.2 Discussion and Outlook

In this last section of the thesis, we first want to briefly discuss the most important short comings of the presented research, for a longer discussion we refer to the conclusions given in Chapters 3 and 4. Further, we want to lay out three different lines of future research that are conceivable taking the results of this thesis as a basis.

With the presented PAI diagnostic, our focus was primarily on an experimental set-up involving a limited number of observations. Future research should aim to analyse more complex (near-) operational set-ups. While the computation of PAI is based on a systematic derivation from LETKF equations, they are computable in a near-operational set-up, irrespective of the quantity and variety of assimilated observations, however the application to larger DA systems might be interesting from the point of view of performance and other techniques that are involved in operational DA. One example for this is that in operational systems ad-hoc methods to inflate the analysis ensemble error covariance are commonly used to prevent overconfidence of the analysis ensemble and potential filter collapse. So far, the proposed PAI diagnostic does not account for inflation techniques and it might be necessary to investigate the influence of inflation methods on the conclusions that can be drawn from PAI diagnostics. Further, the suggested PAI diagnostic is restricted to the analysis influence. An extension of PAI into the forecast is conceivable, similar as in ensemble forecast sensitivity to observations (EFSO).

For the imbalance diagnostic, we suggest to evaluate more cases to gain better understanding of how different synoptic situations affect the generation and evolution of imbalance. In the experiments used to develop the imbalance diagnostics (Chapter 4) inflation methods were active and the results present the effects of the LETKF and the incremental analysis update including inflation methods. In the future it might be interesting to investigate to what extent these inflation methods contribute to the measured imbalance.

Moreover, a synthesis between both developed methods is feasible. Using PAI the imbalance produced by individual or subsets of observations can be computed. The contribution of specific observations to imbalance could shed light on questions related to optimal cycling intervals or localisation settings. To do so, the partial analysis increments have to be propagated into the short-range forecast. As mentioned before, this can be achieved similarly as in EFSO diagnostics, which effectively boils down to replacing the analysis ensemble perturbation matrix in the Kalman Gain with the ensemble perturbation matrix of the forecast, employing a linear model assumption such as is common in EFSO. Nevertheless, there are some open questions that would need to be investigated, such as the treatment of localisation or inflation and its consequences for the computation of propagated PAI. While we think that the effects of localisation in the short-range forecast might be negligible one question is if the propagated PAI or partial imbalance could be computed with reasonable accuracy over a whole cycling interval.

This is especially interesting in the context of the following research idea. Hubans et al. (2022) have recently proposed a method to propagate the observation impact and model uncertainties through multiple DA cycles. While the method of Hubans et al. (2022) involves a variational DA system, an alternate formulation for ensemble DA should be possible (again similar as in EFSO). This means that the contributions of different sources of uncertainty to the ensemble variance could be assessed. Since it is essential to accurately represent the uncertainties in the system and create an ensemble that reflects that, this would yield a third diagnostic tool, which could provide valuable information about the efficiency of a convective-scale DA system.

In conclusion, the methods developed and presented in this thesis — partial analysis increments and measures for imbalance — open the door for further research in the context of convective-scale data assimilation, especially with a focus on optimal use of observations in the data assimilation system. We hope that in future our work will contribute to addressing fundamental questions such as how long does information from a given assimilated observation endure within a convective-scale NWP system, and what impact does it hold for the quality of the forecast.

Bibliography

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble kalman filter for assimilation of radar observations. part i: Storm-scale analyses. *Monthly Weather Review*, **137 (6)**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multicase comparative assessment of the ensemble kalman filter for assimilation of radar observations. part ii: Short-range ensemble forecasts. *Monthly Weather Review*, **138 (4)**, 1273–1292, https://doi.org/10.1175/2009mwr3086.1.

Bachmann, K., C. Keil, G. C. Craig, M. Weissmann, and C. A. Welzbacher, 2020: Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography, and synoptic weather regime. *Monthly Weather Review*, **148 (1)**, 63 – 81, https://doi.org/10.1175/MWR-D-19-0045.1.

Bannister, R. N., 2008: A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, **134 (637)**, 1951–1970, https://doi.org/10.1002/qj.339.

Bauer, P., A. J. Geer, P. Lopez, and D. Salmond, 2010: Direct 4d-var assimilation of all-sky radiances. part i: Implementation. *Quarterly Journal of the Royal Meteorological Society*, **136 (652)**, 1868–1885, https://doi.org/https://doi.org/10.1002/qj.659, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.659.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, https://doi.org/https://doi.org/10.1038/nature14956.

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Monthly Weather Review*, **128 (3)**, 644–667, https://doi.org/10.1175/1520-0493(2000)128<0644:EOSAMF>2.0.CO;2.

Bierdel, L., T. Selz, and G. Craig, 2017: Theoretical aspects of upscale error growth through the mesoscales: an analytical model. *Quarterly Journal of the Royal Meteorological Society*, **143 (709)**, 3048–3059, https://doi.org/10.1002/qj.3160, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3160.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Monthly Weather Review*, **124 (6)**, 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Bouttier, F., and G. Kelly, 2001: 313. Observing systems experiments in the ecmwf 4d-var data assimilation system. ECMWF, Shinfield Park, Reading, URL https://www.ecmwf.int/node/8351, https://doi.org/10.21957/ohibz4sq2.

Buizza, R., 2019: Introduction to the special issue on “25 years of ensemble forecasting”. *Quarterly Journal of the Royal Meteorological Society*, **145 (S1)**, 1–11, https://doi.org/https://doi.org/10.1002/qj.3370, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3370.

Cardinali, C., S. Pezzulli, and E. Andersson, 2004: Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **130 (603)**, 2767–2786, https://doi.org/10.1256/qj.03.205.

Casaretto, G., M. E. Dillon, Y. García Skabar, J. J. Ruiz, and M. Sacco, 2023: Ensemble forecast sensitivity to observations impact (efsoi) applied to a regional data assimilation system over south-eastern south america. *Atmospheric Research*, **295**, 106 996, https://doi.org/https://doi.org/10.1016/j.atmosres.2023.106996.

Craig, G. C., and T. Selz, 2018: Mesoscale dynamical regimes in the midlatitudes. *Geophysical Research Letters*, **45 (1)**, 410–417, https://doi.org/10.1002/2017gl076174.

Daley, R., 1991: *Atmospheric data analysis*. 1, 1 Ed. Cambridge university press.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the*

Royal Meteorological Society, **131 (613)**, 3385–3396, https://doi.org/10.1256/qj.05.108, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1256/qj.05.108.

Diefenbach, T., G. Craig, C. Keil, L. Scheck, and M. Weissmann, 2023: Partial analysis increments as diagnostic for letkf data assimilation systems. *Quarterly Journal of the Royal Meteorological Society*, **149 (752)**, 740–756, https://doi.org/10.1002/qj.4419, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4419.

Dudhia, J., 2014: A history of mesoscale model development. *Asia-Pacific Journal of Atmospheric Sciences*, **50**, 121–131, https://doi.org/10.1007/s13143-014-0031-8.

Errico, R. M., R. Yang, N. C. Privé, K.-S. Tai, R. Todling, M. E. Sienkiewicz, and J. Guo, 2013: Development and validation of observing-system simulation experiments at nasa's global modeling and assimilation office. *Quarterly Journal of the Royal Meteorological Society*, **139 (674)**, 1162–1178, https://doi.org/https://doi.org/10.1002/qj.2027, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2027.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99 (C5)**, 10 143–10 162.

Evensen, G., F. C. Vossepoel, and P. J. van Leeuwen, 2022: *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer Nature.

Eyre, J. R., 2021: Observation impact metrics in nwp: A theoretical study. part i: Optimal systems. *Quarterly Journal of the Royal Meteorological Society*, **147 (739)**, 3180–3200, https://doi.org/https://doi.org/10.1002/qj.4123, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4123.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757, https://doi.org/https://doi.org/10.1002/qj.49712555417, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712555417.

Ge, Y., L. Lei, J. S. Whitaker, and Z.-M. Tan, 2022: The impact of incremental analysis update on regional simulations for typhoons. *Journal of Advances in Modeling Earth Systems*, **14 (10)**, e2022MS003 084.

Geiss, S., L. Scheck, A. de Lozar, and M. Weissmann, 2021: Understanding the model representation of clouds based on visible and infrared satellite observations. *Atmospheric Chemistry and Physics*, **21 (16)**, 12 273–12 290, https://doi.org/10.5194/acp-21-12273-2021.

Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The thorpex observation impact intercomparison experiment. *Monthly Weather Review*, **138 (11)**, 4009–4025, https://doi.org/10.1175/2010MWR3393.1.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble kalman filter localization techniques. *Monthly Weather Review*, **139 (2)**, 511–522, https://doi.org/10.1175/2010mwr3328.1.

Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, **144 (713)**, 1218–1256, https://doi.org/10.1002/qj.3179.

Hamill, T. M., 2006: *Ensemble-based atmospheric data assimilation*, 124–156. Cambridge University Press.

Hirt, M., G. C. Craig, and R. Klein, 2023: Scale interactions between the meso- and synoptic scales and the impact of diabatic heating. *Quarterly Journal of the Royal Meteorological Society*, **149 (753)**, 1319–1334, https://doi.org/10.1002/qj.4456, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4456.

Hohenegger, C., and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *Journal of the Atmospheric Sciences*, **64 (12)**, 4467 – 4478, https://doi.org/10.1175/2007JAS2143.1.

Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017a: Proactive qc: A fully flow-dependent quality control scheme based on efso. *Monthly Weather Review*, **145 (8)**, 3331 – 3354, https://doi.org/10.1175/MWR-D-16-0290.1.

Hotta, D., E. Kalnay, Y. Ota, and T. Miyoshi, 2017b: Efsr: Ensemble forecast sensitivity to observation error covariance. *Monthly Weather Review*, **145 (12)**, 5015 – 5031, https://doi.org/10.1175/MWR-D-17-0122.1.

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3269–3289, https://doi.org/10.1256/qj.05.135, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1256/qj.05.135.

Hu, G., S. L. Dance, R. N. Bannister, H. Chipilski, O. Guillet, B. Macpherson, M. Weissmann, and N. Yussouf, 2022: Progress, challenges and future steps in data assimilation for convection-permitting numerical weather prediction: report on the virtual meeting held on 10 and 12 november 2021. *Atmospheric Science Letters*, **submitted**.

Hu, G., and Coauthors, 2023: On methods for assessment of the influence and impact of observations in convection-permitting numerical weather prediction. 2309.16433.

Hubans, A., L. Berre, Y. Bouteloup, and C. Loo, 2022: Simulation of analysis error contributions arising from uncertainties in the parametrisation of deep convection. *Quarterly Journal of the Royal Meteorological Society*, **148 (748)**, 3099–3114, https://doi.org/https://doi.org/10.1002/qj.4344, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4344.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, **230 (1-2)**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144 (713)**, 1257–1278, https://doi.org/https://doi.org/10.1002/qj.3130, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3130.

Kalnay, E., 2002: Atmospheric modeling, data assimilation and predictability. https://doi.org/10.1017/cbo9780511802270.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble kalman filters. *Tellus A*, **64**, https://doi.org/10.3402/tellusa.v64i0.18462.

Keil, C., F. Baur, K. Bachmann, S. Rasp, L. Schneider, and C. Barthlott, 2019: Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **145 (724)**, 3102–3115, https://doi.org/10.1002/qj.3607, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3607.

Keil, C., F. Heinlein, and G. C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140 (679)**, 480–490, https://doi.org/10.1002/qj.2143, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2143.

Klein, R., 2006: Preface to the special issue on "theoretical developments in tropical meteorology". *Theoretical and Computational Fluid Dynamics*, **20**, 277–278, https://doi.org/10.1007/s00162-006-0035-9.

Klein, R., 2010: Scale-dependent models for atmospheric flows. *Annual review of fluid mechanics*, **42**, 249–274.

Kotsuki, S., K. Kurosawa, and T. Miyoshi, 2019: On the properties of ensemble forecast sensitivity to observations. *Quarterly Journal of the Royal Meteorological Society*, **145 (722)**, 1897–1914, https://doi.org/https://doi.org/10.1002/qj.3534, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3534.

Lange, H., and G. C. Craig, 2014: The impact of data assimilation length scales on analysis and prediction of convective storms. *Monthly Weather Review*, **142 (10)**, 3781–3808, https://doi.org/10.1175/mwr-d-13-00304.1.

Lange, H., G. C. Craig, and T. Janjić, 2017: Characterizing noise and spurious convection in convective data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143 (709)**, 3060–3069, https://doi.org/10.1002/qj.3162.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the nrl atmospheric variational data assimilation adjoint system. *Tellus A*, **56 (3)**, 189–201, https://doi.org/https://doi.org/10.1111/j.1600-0870.2004.00056.x, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0870.2004.00056.x.

Lei, L., D. R. Stauffer, S. E. Haupt, and G. S. Young, 2012: A hybrid nudging-ensemble kalman filter approach to data assimilation. part i: application in the lorenz system.

Tellus A: Dynamic Meteorology and Oceanography, **64 (1)**, 18 484, https://doi.org/10.3402/tellusa.v64i0.18484, https://doi.org/10.3402/tellusa.v64i0.18484.

Lei, L., and J. S. Whitaker, 2016: A four-dimensional incremental analysis update for the ensemble kalman filter. *Monthly Weather Review*, **144 (7)**, 2605–2621, https://doi.org/10.1175/MWR-D-15-0246.1.

Liu, J., E. Kalnay, T. Miyoshi, and C. Cardinali, 2009: Analysis sensitivity calculation in an ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (644)**, 1842–1851, https://doi.org/10.1002/qj.511.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, **20 (2)**, 130 – 141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lynch, P., 2008: The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227 (7)**, 3431–3444, https://doi.org/https://doi.org/10.1016/j.jcp.2007.02.034.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the hirlam model using a digital filter. *Monthly Weather Review*, **120 (6)**, 1019 – 1034, https://doi.org/https://doi.org/10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Majda, A. J., and R. Klein, 2003: Systematic multiscale models for the tropics. *Journal of the Atmospheric Sciences*, **60 (2)**, 393 – 408, https://doi.org/https://doi.org/10.1175/1520-0469(2003)060<0393:SMMFTT>2.0.CO;2.

MetOffice, 2023: Global and uk nwp configurations. Accessed on 5.3.2024, https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/weather-forecasting.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble kalman filter. *Monthly weather review*, **130 (11)**, 2791–2808.

Necker, T., S. Geiss, M. Weissmann, J. Ruiz, T. Miyoshi, and G.-Y. Lien, 2020a: A convective-scale 1,000-member ensemble simulation and potential applications. *Quarterly Journal of the Royal Meteorological Society*, **146 (728)**, 1423–1442, https://doi.org/https://doi.org/10.1002/qj.3744, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3744.

Necker, T., M. Weissmann, Y. Ruckstuhl, J. Anderson, and T. Miyoshi, 2020b: Sampling error correction evaluated using a convective-scale 1000-member ensemble. *Monthly Weather Review*, **148 (3)**, 1229 – 1249, https://doi.org/10.1175/MWR-D-19-0154.1.

Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society*, **144 (714)**, 1667–1680, https://doi.org/https://doi.org/10.1002/qj.3390, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3390.

Okamoto, K., 2017: Evaluation of ir radiance simulation for all-sky assimilation of himawari-8/ahi in a mesoscale nwp system. *Quarterly Journal of the Royal Meteorological Society*, **143 (704)**, 1517–1527, https://doi.org/https://doi.org/10.1002/qj.3022, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3022.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the ncep gfs. *Tellus A: Dynamic Meteorology and Oceanography*, **65 (1)**, 20 038, https://doi.org/10.3402/tellusa.v65i0.20038, https://doi.org/10.3402/tellusa.v65i0.20038.

Palmer, T., 2017: The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *Journal of Advances in Modeling Earth Systems*, **9 (2)**, 730–734, https://doi.org/10.1002/2017MS000999, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017MS000999.

Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the wrf model and its application in the rapid refresh. *Monthly Weather Review*, **144 (1)**, 99–106, https://doi.org/10.1175/MWR-D-15-0219.1.

Piper, D., M. Kunz, F. Ehmele, S. Mohr, B. Mühr, A. Kron, and J. Daniell, 2016: Exceptional sequence of severe thunderstorms and related flash floods in may and june 2016 in germany – part 1: Meteorological background. *Natural Hazards and Earth System Sciences*, **16 (12)**, 2835–2850, https://doi.org/10.5194/nhess-16-2835-2016.

Potthast, R., 2019: Documentation of the data assimilation coding environment (dace).

Puh, M., C. Keil, C. Gebhardt, C. Marsigli, M. Hirt, F. Jakub, and G. C. Craig, 2023: Physically based stochastic perturbations improve a high-resolution forecast of convection. *Quarterly Journal of the Royal Meteorological Society*, **n/a (n/a)**, https://doi.org/https://doi.org/10.1002/qj.4574, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4574.

Reinert, D., and Coauthors, 2020: Dwd database reference for the global and regional icon and icon-eps forecasting system. Accessed on 19.3.2024, https://www.dwd.de/DWD/forschung/nwv/fepub/icon_database_main.pdf.

Richardson, L. F., 1922: *Weather prediction by numerical process*. University Press, URL https://archive.org/details/weatherpredictio00richrich/weatherpredictio00richrich/page/n11/mode/2up.

Scheck, L., P. Frerebeau, R. Buras-Schnell, and B. Mayer, 2016: A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **175**, 54–67, https://doi.org/https://doi.org/10.1016/j.jqsrt.2016.02.008.

Scheck, L., M. Weissmann, and L. Bach, 2020: Assimilating visible satellite images for convective-scale numerical weather prediction: a case-study. *Quarterly Journal of the Royal Meteorological Society*, **nil (nil)**, qj.3840, https://doi.org/10.1002/qj.3840.

Scheck, L., M. Weissmann, and B. Mayer, 2018: Efficient methods to account for cloud-top inclination and cloud overlap in synthetic visible satellite images. *Journal of Atmospheric and Oceanic Technology*, **35 (3)**, 665 – 685, https://doi.org/10.1175/JTECH-D-17-0057.1.

Schomburg, A., C. Schraff, and R. Potthast, 2015: A concept for the assimilation of satellite cloud information in an ensemble kalman filter: single-observation experiments. *Quarterly Journal of the Royal Meteorological Society*, **141 (688)**, 893–908, https://doi.org/https://doi.org/10.1002/qj.2407, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2407.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periáñez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Quarterly Journal of the Royal Meteorological Society*, **142 (696)**, 1453–1472, https://doi.org/10.1002/qj.2748.

Schroettle, J., M. Weissmann, L. Scheck, and A. Hutt, 2020: Assimilating visible and infrared radiances in idealized simulations of deep convection. *Monthly Weather Review*, **148 (11)**, 4357 – 4375, https://doi.org/10.1175/MWR-D-20-0002.1.

Selz, T., M. Riemer, and G. C. Craig, 2022: The transition from practical to intrinsic predictability of midlatitude weather. *Journal of the Atmospheric Sciences*, **79 (8)**, 2013 – 2030, https://doi.org/10.1175/JAS-D-21-0271.1.

Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **140 (685)**, 2672–2679, https://doi.org/10.1002/qj.2343.

Sommer, M., and M. Weissmann, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus A: Dynamic Meteorology and Oceanography*, **68 (1)**, 27 885, https://doi.org/10.3402/tellusa.v68.27885, https://doi.org/10.3402/tellusa.v68.27885.

Stephan, K., S. Klink, and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model cosmo-de at dwd. *Quarterly Journal of the Royal Meteorological Society*, **134 (634)**, 1315–1326, https://doi.org/10.1002/qj.269, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.269.

Stratman, D. R., and K. A. Brewster, 2017: Sensitivities of 1-km forecasts of 24 may 2011 tornadic supercells to microphysics parameterizations. *Monthly Weather Review*, **145 (7)**, 2697 – 2721, https://doi.org/10.1175/MWR-D-16-0282.1.

Talagrand, O., 1997: Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, **75 (1B)**, 191–209, https://doi.org/10.2151/jmsj1965.75.1B_191.

Tribbia, J., 2020: *Normal Mode Functions and Initialization*, 63–78. Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-60963-4_2, URL https://doi.org/10.1007/978-3-030-60963-4_2.

Vallis, G. K., 2017: *Equations of Motion*, 3–54. 2nd ed., Cambridge University Press, https://doi.org/10.1017/9781107588417.002.

Vetra-Carvalho, S., M. Dixon, S. Migliorini, N. K. Nichols, and S. P. Ballard, 2012: Breakdown of hydrostatic balance at convective scales in the forecast errors in the met office unified model. *Quarterly Journal of the Royal Meteorological Society*, **138 (668)**, 1709–1720.

Zeng, Y., A. de Lozar, T. Janjic, and A. Seifert, 2021: Applying a new integrated mass-flux adjustment filter in rapid update cycling of convective-scale data assimilation for the cosmo model (v5.07). *Geoscientific Model Development*, **14 (3)**, 1295–1307, https://doi.org/10.5194/gmd-14-1295-2021.

Zhang, F., M. Minamide, and E. E. Clothiaux, 2016: Potential impacts of assimilating all-sky infrared satellite radiances from goes-r on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, **43 (6)**, 2954–2963, https://doi.org/https://doi.org/10.1002/2016GL068468, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL068468.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, **141 (687)**, 563–579, https://doi.org/https://doi.org/10.1002/qj.2378, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2378.

List of Figures

List of Abbreviations

AMDAR Aircraft Meteorological Data Relay.

COSMO COnsortium for Small-scale MOdeling.

DA Data Assimilation.

DPSDT Surface Pressure Tendency.

DWD Deutscher Wetterdienst.

EFSO Ensemble Forecast Sensitivity to Observations.

ETKF Ensemble Transform Kalman Filter.

FSO Forecast Sensitivity to Observations.

IAU Incremental Analysis Update.

ICON ICosahedral nONhydrostatic.

KENDA Kilometer Scale ENsemble Data Assimilation.

LETKF Local Ensemble Transform Kalman Filter.

LHN Latent Heat Nudging.

LOC Localisation.

NWP Numerical Weather Prediction.

NWV Numerische Wettervorhersage (NWP).

OSE Observation System Experiment.

OSSE Observation System Simulation Experiment.

PAI Partial Analysis Increment.

RASO Radiosonde Observations.

REFL Reflectance.

RH Relative Humidity.

SWE Shallow Water Equations.

SYNOP Surface Synoptic Observations.

UTC Coordinated Universal Time.

VIS Visible Satellite Observations.

VMD Vertical Motion Diagnostic.

WTG Weak Temperature Gradient.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor George Craig, who has been a constant source of inspiration and motivation. Your ability to clearly articulate complex concepts provided many eye-opening moments during our discussions. You created an atmosphere in the group that made me feel truly welcome and reassured that I was in the right place. Throughout my PhD, I have felt that this group is an example of how research should be done and I am deeply grateful to have been a part of it. Thank you for everything!

I am equally grateful to Professor Martin Weissmann, thank you for giving me the opportunity to work on this topic and sharing your ideas with me. I have learned so much from you, and our conversations about science and beyond have always been enlightening and enjoyable.

Leonhard Scheck, thank you for sharing your competence and providing support whenever I needed it. You have assisted this thesis in numerous ways, and our scientific discussions have always been insightful and fun. It was a pleasure to share the office with you for such a long time.

Christian Keil, I am grateful for all your good advice along the way, your guidance has been greatly appreciated.

Tobias Selz, I truly believe you are a world-leading expert. Thank you for answering my questions, sharing your knowledge, and offering me a temporary office space. Your support has been crucial.

Barbara, thank you for all your help and for bringing so much positive energy to the group. Your warmth and approachable nature have made a significant difference.

To all my PhD colleagues, postdocs, and former colleagues, including Federico, Flo, Jonas, Kirsten, Matjaz, Mirjam, Oriol, Philipp, Robert, Stefan, Stefano, Takumi, Yvonne, despite the tough start with Covid sending us all into home office, I have thoroughly enjoyed our time together. The collegial atmosphere, the countless stews and M13s and the productive exchange of knowledge were a lot of fun and incredibly valuable.

To my colleagues in Vienna, Tobias and Philipp, thank you for the fruitful discussions that have greatly contributed to my research.

Anne, thank you for always providing advice and proofreading parts of this thesis. Our shared interest in meteorology brought us together during our master's thesis, and your example has been a source of motivation for me. I am truly grateful to call you my friend.

To my parents, thank you for always having my back. Your understanding and encouragement have helped me pursue my goals, just as you pursued yours. Thank you for providing unwavering support.

Lisi, the way you live your life inspires me every day. You are the strongest person I know!

Max, from our school days, you taught me the importance of asking the right questions when seeking help. Your influence has been long-lasting and deeply appreciated.

Leon, thank you for all your help regarding this thesis, whether it was proofreading, motivating me, listening to my problems, asking questions about my work, or providing much-needed distractions and laughter. It means the world to me that I can always count on your support.