MASTERARBEIT | MASTER'S THESIS

Titel | Title
Scale-dependent assimilation of cloud fraction

verfasst von | submitted by
Andrea Hochebner, BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree
of
Master of Science (MSc)

Wien, | Vienna, 2024

Studienkennzahl lt. Studienblatt | degree programme code as it appears on the student record sheet:	UA 066 614
Studienrichtung lt. Studienblatt | degree programme as it appears on the student record sheet:	Masterstudium Meteorology
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Martin Weißmann
Mitbetreut von | Co-Supervisor:	Dott. Dott. ric. Stefano Serafin

Abstract

Predicting and modelling convective-scale features, such as convective cells and associated hazards, provides many challenges for numerical weather prediction (NWP) models. The convective scale is sensitive to small-scale errors, limited by intrinsic predictability and not easily constrained by conventional observations, which also poses a challenge for data assimilation (DA). Visible reflectance observations offer the necessary temporal and spatial resolution. Still, they are underused in operational centres because the nonlinear relation between the state and the observed variable violates fundamental assumptions of many DA algorithms: the linearity of the observation operator and Gaussian error distributions. This thesis proposes a new form of assimilating visible satellite information by computing a cloud fraction with a threshold over a changeable area. Like superobbing, pixel-by-pixel assimilation is avoided, and, especially when more than one scale of cloud fraction is assimilated, nonlinear effects are possibly mitigated. The relative impact of single and multiscale cloud fraction assimilation is assessed with idealised observing system simulation experiments (OSSEs), where the truth is perfectly known and systematic operator or model deficiencies are neglected. The conducted OSSE comprise the Weather Research and Forecast (WRF) model for forecasting, the Data Assimilation Research Testbed (DART) and its implementation of the Ensemble Adjustment Kalman Filter (EAKF) to assimilate observations and the RTTOV-MFASIS forward operator to simulate those observations. The cloud fractions are precomputed from visible reflectance images and assimilated with an empirically estimated observation error. All conducted experiments have been evaluated with a probabilistic formulation of the fraction skill score: the pFSS. Experiments assimilating a single scale of cloud fraction computed over 16x16 km exhibit improved skill compared to a direct visible reflectance benchmark experiment on the same scale for cloud fraction, precipitation, and visible reflectance. Assimilating all available cloud fraction scales, ranging from 256x256 km to 16x16 km scales, does improve pFSS scores but leads to spurious convection and adds a bias to the humidity profile with each assimilation cycle. These issues have been alleviated by updating coarse scales less frequently and assigning a better-suited observation error estimate, leading to improvements of up to 20 % pFSS compared to the free ensemble for visible reflectance during the free forecast and a consistently positive impact on all scales for predicted precipitation.

Kurzfassung

Die Vorhersage und Modellierung von konvektiven Zellen und damit verbundener Gefahren stellt numerische Wettervorhersagemodelle (NWP) vor viele Herausforderungen. Die konvektionserlaubende Skala reagiert empfindlich auf kleine Fehler, ist durch die intrinsische Vorhersagbarkeit begrenzt und lässt sich oft nicht durch konventionelle Beobachtungen einschränken, was auch eine Herausforderung für die Datenassimilation (DA) darstellt. Satellitenbeobachtungen im sichtbaren Bereich (VIS) besitzen die notwendige zeitliche und räumliche Auflösung, werden jedoch operationell selten genutzt, da die nicht-lineare Beziehung zwischen dem Modelzustand und der beobachteten Variable grundlegende Annahmen vieler DA-Algorithmen verletzt: die Linearität des Beobachtungsoperators sowie Gaußsche Fehlerverteilungen. Diese Arbeit präsentiert eine neue Form der Assimilation von VIS-Bildern, die sogenannte cloud fraction, welche mit einem Schwellenwert über einem veränderlichen Gebiet berechnet wird. Ähnlich wie mit Superobbing wird eine pixelweise Assimilation vermieden, und das Assimilieren von multiplen Skalen kann potenziell helfen, nicht-lineare Effekte zu vermindern. Die relativen Auswirkungen von cloud fraction Assimilationen von einzelnen oder mehreren Skalen werden mit idealisierten Beobachtungssystemsimulationsexperimenten (OSSEs) evaluiert. Das experimentelle Design besteht aus dem Weather Research and Forecast (WRF)-Modell für Vorhersagen, sowie dem Data Assimilation Research Testbed (DART) mit dem Ensemble Adjustment Kalman Filter (EAKF), um synthetisch erstellte cloud fraction Beobachtungen zu assimilieren. Es wurde ein empirisch erstelltes Beobachtungsfehlermodell angewendet und alle durchgeführten Experimente wurden mit einer probabilistischen Variante des Fraction Skill Score (pFSS) evaluiert. Experimente, die 16x16 km cloud fraction assimilierten, weisen im Vergleich zu dem Benchmark-Experiment mit direkten VIS-Beobachtungen Verbesserungen für cloud fraction, Niederschlag und VIS auf. Die Assimilation aller verfügbaren cloud fraction Skalen, die Beobachtungen von 16x16 km bis 256x256 km umfasst, verbessert zwar die pFSS-Werte, führt jedoch zu unerwünschter, zufällig verteilter Konvektion und einer Verzerrung des Feuchteprofils. Diese Probleme wurden durch die seltenere Assimilierung von großflächigen cloud fraction Beobachtungen und Verbesserungen in der Beobachtungsfehlerschätzung gemildert. Insgesamt konnten mehrskalige cloud fraction Assimilationsexperimente den pFSS für VIS-Vorhersagen bis zu 20 % verbessern im Vergleich zum freien Ensemble und eine positive Auswirkung auf allen Skalen für vorhergesagten Niederschlag konnte ebenfalls festgestellt werden.

Contents

1 Introduction and motivation

With advancements in the last decades, numerical weather prediction (NWP) models on a global scale allow for precise forecasting of synoptic-scale events. However, predicting convective-scale features, such as convective cells and associated hazards, provides many challenges for models, and there is still limited success in forecasting convection-permitting scales (Clark et al., 2016). Processes that are often not accurately represented, such as low stratus and convective precipitation, can significantly impact society. Deep convective systems such as thunderstorms and their associated rainfall threaten the public and require timely and accurate warnings (Scheck et al., 2020). Deficiencies in the model representation of clouds also contribute strongly to the uncertainty in solar power production forecasts (Kurzrock et al., 2018). Clouds are the first manifestation of convective processes and are detectable even before precipitation forms. The prediction of convection would strongly benefit from better positioning and constraining clouds within NWP systems.

The convective scale contains various meteorological features ranging from 4-40 km. The effective resolution of a numerical model is 5-10 times larger than the applied grid spacing (Skamarock, 2004). In contrast to global models, the limited domain size of regional models allows for a convection-permitting grid resolution of 1-2 km, which can resolve convective-scale features at least partially. As the temporal predictability of those models ranges from hours to a few days (Zhang et al., 2019), the main goal of convective-scale NWP is to use the large-scale boundary conditions provided by global models to forecast the position and numbers of individual convective cells.

However, a sufficient numerical grid resolution does not ensure successful regional forecasts. Partly, this is due to the intermittent and stochastic nature of convection, the predictability of which is intrinsically limited (Bachmann et al., 2020). Although high-resolution models can partly resolve moist convection explicitly, their simulations still suffer from several sources of systematic errors. How a model represents clouds is influenced by multiple parameterization schemes containing many parameters that often are not well constrained (Geiss et al., 2021).

Another forecast error source stems from the uncertainty of the initial conditions. The better the initial state of the atmosphere is known and represented in the model, the better the subsequent forecast. Data assimilation (DA) aims to reduce the uncertainty of the initial conditions by combining short-term forecasts with observations. The obtained analysis can serve as enhanced initial conditions for the forecasting. Thus, increasing the resolution of numerical models represents a challenge for DA systems, as the assimilated information should constrain the model state on all its resolved scales.

Observations constraining the convective scale are relatively scarce. Ground-based, conventional observation types usually have a temporal and spatial inhomogeneous distribution and are often not representative of a larger area. Indirect remote-sensing observations like radar can monitor larger neighbourhoods and constrain the thermodynamic cloud structure and height. However, geographic shadowing and gaps in the radar network limit data quality. Additionally, this observation type needs precipitating particles to measure the reflected microwave radiation and, therefore, already formed, larger hydrometeors (Fabry & Meunier, 2020). Satellite observations, however, cover large parts of Earth's surface and offer a high observation frequency (Kugler et al., 2023), which is especially important in less observation-dense parts of the world (Bauer et al., 2015). Assimilated satellite products, as listed by Eyre et al. (2022), include passive temperature or humidity soundings, infrared (IR) imagery, microwave (MW) imagery and visible (VIS) imagery.

Satellite observations for convective-scale modelling

Observations stemming from LEO[1] satellites, such as MW imagery, play an important role for global NWP DA systems (Geer et al., 2018). LEO satellites only observe an area once per day, which significantly limits the temporal resolution. Instruments on geostationary satellites, though, can provide temporally and spatially high-resolved data of VIS and IR observations suitable for regional forecasting. IR channels from geostationary orbits observe the humidity field and can provide valuable information regarding convective initiation. However, their signal is significantly changed in the presence of clouds, as IR radiances are sensitive to water vapour and cloud height and, therefore, cannot detect low-level clouds. On the other hand, 0.6 μm visible reflectance observations use the visible spectrum during the daytime, meaning that the observed radiation from any vertical level is not attenuated, and the observation is height-independent. Visible satellite channels can, therefore, provide complementary information to IR observations about cloud distributions and properties, as they allow for a clear distinction between low clouds and ground (Scheck et al., 2016).

Cloud-affected satellite observations remain still underused in DA systems, as their assimilation can violate assumptions of currently used DA schemes and may lead to a suboptimal forecast update (Geer et al., 2018; Gustafsson et al., 2018). The data retrieved by satellites is usually not a model variable, but a reflectance or irradiance measurement, and a suitable forward operator h is needed to create the model equivalents to update the model state with the observed information. Many DA algorithms assume a linear h, which also applies to the sequential Kalman Filter used in this study, the EAKF (Anderson, 2001). However, cloud-affected satellite observations are not linearly related to the model state variables, and their error distributions are often non-Gaussian (Kugler et al., 2023).

Another issue that arises as a prerequisite for assimilating cloud-related observations is the need for a good representation of clouds in the model. Assimilated and simulated clouds should follow a similar climatology, which is often unfulfilled due to systematic model representation errors (Geiss et al., 2021; Scheck et al., 2020). Deficiencies in representation can introduce a bias to the first guess (FG) forecast and violate the basic

---

[1] LEO - Low Earth Orbits

assumption of an unbiased estimate in currently used DA schemes (Gustafsson et al., 2018).

Therefore, operational DA setups have frequently excluded cloud-affected satellite channels or assimilated only retrieved information. Currently, different all-sky assimilation approaches using water vapour channels from LEO satellites are operational in centres around the world (Geer et al., 2018) and have been studied in many instances (Cintineo et al., 2016; Jones et al., 2014; Okamoto, 2017). Regarding regional modelling, IR observations are still not operationally used due to the reasons above (Geer et al., 2018).

However, the first assimilations of visible reflectances were made possibly recently with the development of a suitable and fast enough forward operator: RTTOV-MFASIS (Saunders et al., 2018). It is a lookup table-based, radiative transfer model for simulating top-of-atmosphere radiances, first introduced by (Scheck et al., 2016) and soon added to RTTOV (Radiative Transfer for TOVS) (Saunders et al., 2018). It simulates the top of atmosphere reflectance as seen by the satellite with a function depending on particle sizes for water and ice clouds, as well as vertically integrated optical depths. It also takes the surface albedo, the sun, the satellite zenith angle and the scattering angle into account (Scheck et al., 2016). Scheck et al. (2020) and Schröttle et al. (2020) have explored the potential of direct reflectance assimilation and built the foundation for its operational implementation in the regional model of DWD[2] (Schomburg & Bach, 2023). As the impact of direct visible reflectance has only been investigated for the NWP and DA system of DWD, Kugler et al. (2023) explored direct visible reflectance assimilation with the Weather Research and Forecast (WRF) model with promising results. The motivation of this thesis is to investigate the assimilation of visible reflectance information further and contribute to a better understanding of its impact.

Pixel-by-pixel assimilations of reflectances are avoided as observation errors are assumed to be uncorrelated (Kugler et al., 2023), as this is not the case with neighbouring pixels of a satellite image. Pixel-based verification methods, such as the root mean squared error (RMSE), have been proven in many instances to be suboptimal for verification purposes (Roebber et al., 2004; Rossa et al., 2008), as they penalize a location or intensity mismatch twice. Neighbourhood-based metrics like the Fraction Skill Score (FSS) introduced by Roberts and Lean (2008) are less sensitive to minor displacement errors and therefore favoured for spatial verification (Necker et al., 2023). A more or less equivalent approach for the assimilation of satellite images is the so-called superobbing. Averaging over radiances in some way to generate superobs acts similar to a low-pass filter, helps to smooth the images of the departures and reduces noise (Duncan et al., 2023). Superobbing for VIS reflectance was shown to be beneficial for the forecast by Scheck et al. (2020) for the NWP and DA system at DWD. However, Kugler et al. (2023) did not find a difference between assimilating thinned visible reflectance observations and superobbed ones, which might be due to superobbing algorithms in WRF being optimized for different observation types (Sun & Wang, 2013). Therefore, this thesis proposes a new type of superobbing: cloud fraction.

---

[2] DWD - German Weather service

Cloud fraction

Assimilating cloudiness in relatively large areas, instead of considering the pixel scale of the initial visible reflectance image, might prove beneficial to the state update, as it reduces the double penalty error from displacements of clouds. Instead of simply averaging, the VIS image is transformed into a binary field with a threshold before it is averaged over the computation area. This kind of superobbing for VIS observations is called cloud fraction. A high threshold, for example, 0.6 μm VIS reflectance, helps to isolate water clouds and, therefore, clear signals of convective activity from high ice clouds (Scheck et al., 2020). This is important, as VIS imagery is height-independent and cannot distinguish between ice and water clouds. However, this also means that there is no introduced bias from height-dependency of the observation like with IR imagery. Compared to a direct VIS assimilation approach, ensemble members with similar cloud fraction values as the observation will be given more weight during a cloud fraction assimilation, even though the placement of small convective systems inside the observed area might differ. Ensemble members do not need to model the convective system perfectly, which is quite difficult to achieve given the stochastic nature of convection. Even if the exact placement of each individual cell might not be feasible, having the same amount of convective activity in a domain can benefit cloudiness and precipitation forecasts. The potential of assimilating such structural information with a threshold-based approach from VIS images has not yet been explored and will be the main focus of this thesis.

Cloud fraction can be evaluated on areas of arbitrary size, enabling a multiscale assimilation approach. Going from larger scales to finer ones might effectively alleviate non-linearity, as the modelled overall convective activity is adjusted to the observed one at each scale. A similar concept of gradually adjusting an ensemble to the observations has been investigated by Ying (2019) and Ying et al. (2023), with good results in enhancing the position of an extensive convective system on a large scale. A more challenging scenario for DA algorithms on the convective scale is one with deep convection triggered at random locations. It exhibits low predictability, as it is susceptible to initial conditions and has to account for fast-growing errors and interacting convective cells (Bachmann et al., 2020). This work will explore the possibilities of assimilating multiple scales of cloud fraction observations in this highly unpredictable random case.

1.1 Research questions and outline

The present work aims to provide insight into two main research topics:

I Is the impact of cloud fraction assimilation positive and how does it compare with that of direct visible reflectance assimilation?

II Is it useful to assimilate cloud fractions at different spatial scales, and if yes, how often should the different scales be assimilated?

The first research question focuses on single-scale cloud fraction assimilation and its possible impact on forecasts of precipitation and cloudiness. This study uses perfect model OSSEs (Observing System Observation experiments) in an idealized experimental design to evaluate the relative impact of cloud fraction assimilation. An identical assimilation cycling is also performed with direct VIS reflectance assimilation, which serves as a benchmark experiment. As there is no observation error model for cloud fractions yet, an empirical error model is obtained by transposing the instrument error applied in Kugler et al. (2023) for VIS to cloud fractions. Results are evaluated with a probabilistic formulation of the Fraction Skill Score, the pFSS (Necker et al., 2023).

The second research topic relates to the impact of multiscale cloud fraction, where OSSEs are conducted by assimilating cloud fractions on coarser and finer scales at the same assimilation cycling step. To ensure independent observations, only the differences of the observations on different scales are assimilated. As finer scales exhibit a higher temporal volatility than coarser ones, experiments featuring a different assimilation frequency for different scales have also been conducted.

The context of this thesis will be provided in chapter 2, which gives a brief overview of mesoscale convection, numerical weather prediction and data assimilation. It also describes the data assimilation algorithm applied in this study and the verification method pFSS.

Chapter 3 focuses on the experimental design and its components. It describes the technical aspects of the conducted OSSEs and gives detailed information on how cloud fractions are calculated from a satellite image. Section 3.4 illustrates the empirically obtained observation error variance models used in this study.

The results of all conducted cloud fraction assimilation experiments are shown in chapter 4. Potential issues with the multiscale approach are highlighted in section 4.2.1, and an experiment featuring an enhanced error model alleviating those issues is introduced in section 4.2.2.

The last chapter, 5, is dedicated to discussing said results and offers conclusions and suggestions for further investigations.

2 Background and methods

The following chapter describes the theoretical and technical background for this study as well as the verification methods.

2.1 Mesoscale convection

Convection generally refers to mass transport in the atmosphere in an upward direction.

Definition of the mesoscale

Motion in the atmosphere has a broad spectrum of temporal and spatial scales. The timescales range from under a second for atmospheric turbulence to weeks for planetary-scale baroclinic waves. The spatial scales of motion also range from microphysical processes below one centimetre to thousands of kilometres (Markowski & Richardson, 2010), as is illustrated by Figure 2.1.

Over a dozen different length scale limits for the mesoscale have been proposed. Orlanski (1975)'s definition with a mesoscale process spanning 2 to 2000 km is one of the most widely used (Markowski & Richardson, 2010) and will also be used in this thesis. Even though convective clouds, precipitation, and associated phenomena have characteristic temporal and spatial scales that can extend below the lower range of the mesoscale and can be forced by processes at or above the upper end, cumulus and cumulonimbus convection, such as thunderstorms, are mostly mesoscale phenomena (Trapp, 2013). Convection-permitting NWP models, as in this thesis, have a spatial and temporal resolution that is high enough to partly resolve mesoscale convection.

Convective storms

This section follows closely McMurdie and Houze (2006).

Thunderstorms (or convective storms) are forms of local deep moist convection associated with various severe weather conditions. One can differentiate between:

- Single-cell storms: an individual convective cell with a single up- and downdraft that is also called an ordinary thunderstorm;
- Multicell storms: organized groups of more than one convective cell;
- Supercell storms: intense and long-lived clusters of cells characterized by rotating updrafts;

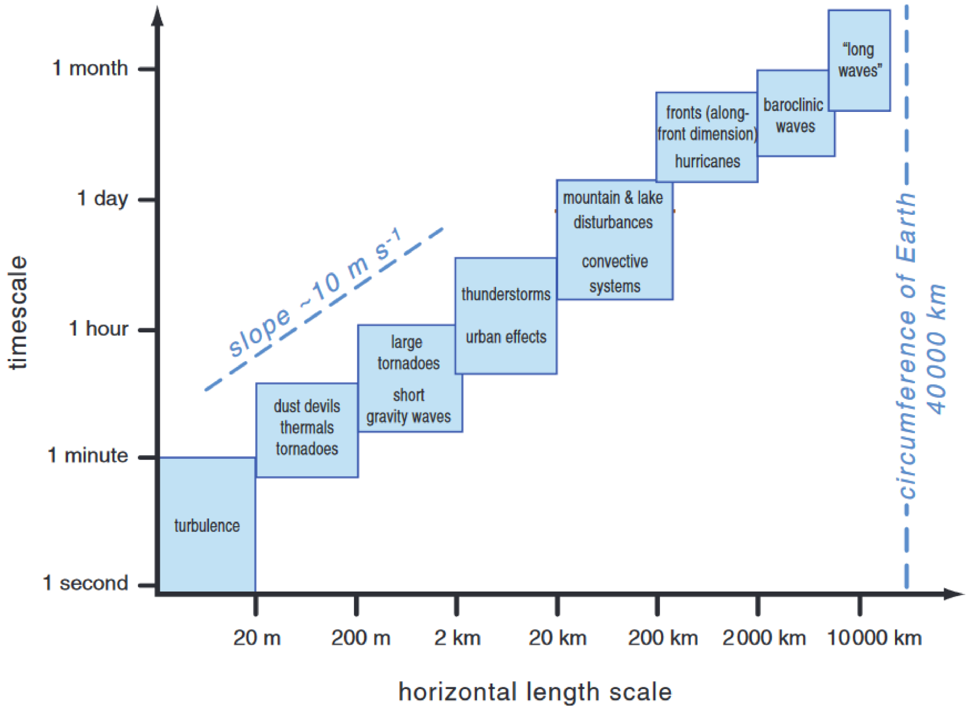


Figure 2.1: Different atmospheric processes with their characteristic time and spatial scales. Figure adapted from Markowski and Richardson (2010).

Ordinary thunderstorms

Ordinary thunderstorms are local convective systems in a small, unstable air mass and form a small, isolated cumulonimbus cloud. These systems generally develop just one main precipitation shower and are also called single-cell storms. Their formation is associated with weak vertical wind shear. The typical lifespan of a single-cell storm is around 30 minutes and contains three main stages. The first is the developing stage or cumulus stage without significant precipitation. During the mature stage, as the system grows, formed precipitation particles can begin to fall below the cloud base, and a downdraft circulation is initiated. In the dissipation stage, as the unsaturated air below the cloud base is cooled, the updraft loses a source of warm and moist air and blocks the growth of more rain droplets — leading to the demise of the storm.

2.2 Numerical weather predication

This section is based on Pu and Kalnay (2018) and Trapp (2013) and attempts to give a short overview of numerical weather prediction.

Advancements in technology have enabled numerical weather prediction (NWP) to become a central component of today's weather forecasts. The basic concept is that a set of partial differential equations (PDEs) governs the evolution of the atmosphere. NWP models integrate these PDEs numerically forward in time to predict the future atmospheric state. Therefore, the initial atmospheric conditions must be described as accurately as possible.

Primitive equations

The motion and evolution of the atmospheric state are captured in seven basic equations: conservation of mass (continuity equation), conservation of energy, conservation of water mass, the state equation for an ideal gas and one equation for conservation of momentum for each velocity component. Depending on the design of the microphysics scheme, the conservation of water is split into equations for cloud water, rainwater and water vapour, and in the case of a mixed-phase scheme, equations for snow, graupel/hail and cloud ice are added.

Those universal equations are often solved in approximate form. For example, the vertical momentum equation is simplified on a large-scale NWP with the hydrostatic balance assumption.

Numerical frameworks and grid staggering

There is generally no analytical solution to the aforementioned PDEs except for several approximated cases. Therefore, the equations are discretized with spectral or finite volume methods to find approximate solutions. A numerical scheme needs to be stable and consistent for the numerical solution to converge with the true one. NWP models today use semi-Lagrangian schemes for spatial discretization on global scales, finite volume schemes for temporal discretization, or high-order Runge-Kutta schemes for both. Another degree of freedom is the arrangement of model variables on different grid points, with models usually adapting a staggered grid approach. A staggered grid combines nodal points on different positions to achieve a more accurate characterization of PDEs on the discrete grid, both in the horizontal and in the vertical direction.

Effective resolution

Regarding the resolution of an NWP system, one has to mention the difference between numerical resolution and resolved features. Numerical dissipation is introduced either by the numerical scheme itself or added on purpose to avoid the onset of numerical instability. Therefore, the effective resolution of the model is always below the theoretically possible maximal value of twice the grid spacing. Most NWP models resolve effectively 5 to 10 times their grid spacing, and any processes on smaller scales are not resolved due to the numerical inaccuracy (Abdalla et al., 2013; Skamarock, 2004).

Parametrizations and sub-grid processes

Figure 2.1 shows that atmospheric motion covers various scales. As the PDEs are numerically discretized, the grid resolution of the model is always limited. Any sub-grid-scale process cannot be explicitly resolved, but its contribution cannot be ignored. Regional-scale, convection-permitting NWP models resolve deep convective processes explicitly, but turbulent heat and moisture fluxes, condensation, evaporation and friction are below the resolution. Therefore, parameterization schemes are necessary to properly account for unresolved processes' impact and interactions with larger, resolved scales: the averaged net effect of the sub-grid mechanism must be formulated and applied to the model grid. Precipitation, for example, is computed by the microphysics parametrization scheme and highly depends on its implementation.

2.3 Data assimilation

NWP is an initial/boundary value problem: an estimate of the present state of the atmosphere is necessary to predict its future evolution. The better the initial conditions are represented, the better the quality of the forecast will be. Today's operational weather centres use a statistical combination of observations and short-range forecasts, so-called data assimilation (DA) (Kalnay, 2002). As observations often do not measure model variables directly and exhibit a nonuniform distribution in time and space, the observed information is combined with a 'background' defined on all model grid points. The best background available is usually a short-range forecast, called first-guess. It is then combined with observations valid at that time to yield an analysis, which may serve as the initial conditions for the NWP model, producing a first guess again and long-term forecasts (Kalnay, 2002). This concept, known as cycling, is depicted in Figure 2.2 for a regional model.

When combining observed and background information, one needs to determine how much weight is given to each ingredient. The simplest realization of the concept is the 1-D Bayes Theorem:

$$P(T|T_o, C) = \frac{P(T_o|T, C) \cdot P(T|C)}{P(T_o|C)}, \tag{2.1}$$

where $P(T|T_o, C)$ is the probability of the event T, given the observed event T_o and prior information C. The numerator in Equation 2.1 is the likelihood of observing T_o, assuming T is the true value and taking C into account, multiplied with the prior $P(T|C)$. The denominator $P(T_o|C)$ serves as a normalization. An often-made fundamental assumption in data assimilation is that the probabilities P for the prior and likelihood are Gaussian and can be described by a mean value $\overline{T}$ and its standard deviation σ_T - resulting in a Gaussian posterior (Evensen et al., 2022).

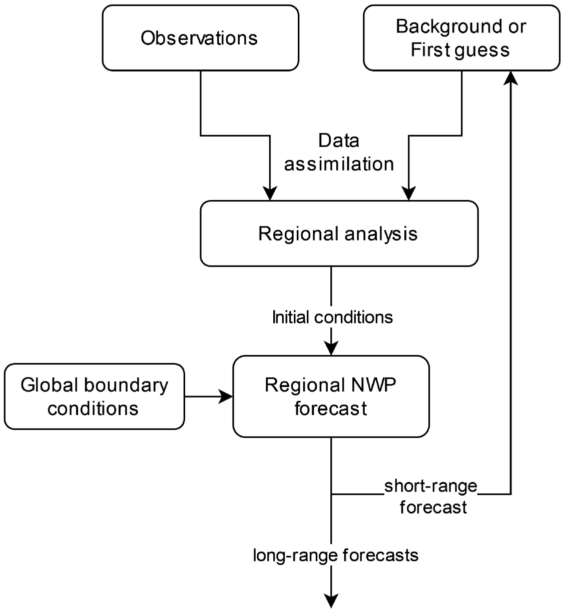


Figure 2.2: Schematic regional NWP analysis cycle: Observations valid at the same time as the background information are combined to produce an analysis for the regional NWP model, which can then predict a first-guess and, if needed, long-term forecasts. The regional model requires boundary conditions, for example provided by a global NWP model. Figure adapted from Kalnay (2002).

2.3.1 Ensemble data assimilation and the EAKF

Ensemble data assimilation

As the atmosphere is a nonlinear system, even small perturbations of the initial conditions can lead to vast differences if forwarded in time, which fundamentally limits the predictive skill of NWP (Bauer et al., 2015). One way to attempt to capture how the uncertainty of the analysis and the forecast evolves is ensemble prediction: a Monte Carlo method for propagating the error statistics. Starting with an ensemble of samples, each ensemble member is integrated forward in time by the model separately to create a number of model realizations that ideally represent the uncertainties of the forecast (Evensen et al., 2022). The spread of the ensemble can quantify forecast uncertainty but is limited by the finite sample size (Bauer et al., 2015).

The Ensemble Adjustment Kalman Filter - EAKF

DA algorithms also need to be computationally feasible. The matrix of the model state contains all model variables and is around 10^8x10^8 for regional models, depending on the grid spacing. The observation state contains all observed quantities and is typically smaller than the model state, as its size depends on the used observation types. To avoid dealing with those huge matrices, a sequential approach assimilating one observation after another has been proposed by Anderson (2001). The following aims to give an overview of the so-called Ensemble Adjustment Kalman Filter (EAKF) based on Karspeck and Anderson (2007) and a tutorial created by NCAR[1] (NCAR, 2009).

The EAKF assumes independent observation errors and, therefore, error distributions, which allows the assimilation of subsets of observations sequentially. This leads to the computation of many small scalar problems instead of dealing with huge matrices. It can be summarized in three main steps, repeated for each assimilated observation.

Firstly, we assume a single observation y^o and a prior ensemble with N members. Each member i has a state vector x_i^b that can be transformed to observation space with an observation operator h, resulting in model equivalents y_i^b: $y_i^b = h(x_i^b)$. Overlines indicate the mean over the members with the variances $\sigma^2_{x^b}$ and $\sigma^2_{y^b}$, respectively. The purpose of the algorithm is to compute the mean $\overline{x^a}$ and standard deviation σ_{x^a} of the posterior.

Step A.

In step A., the mean of the posterior $\overline{y^a}$ and variance $\sigma^2_{y^a}$ in observation space are computed:

$$\overline{y^a} = \sigma^2_{y^a}\left(\frac{\overline{y^b}}{\sigma^2_{y^b}} + \frac{y^o}{\sigma^2_{y^o}}\right) \qquad \text{with } \sigma_{y^a} = \left(\frac{1}{\sigma^2_{y^o}} + \frac{1}{\sigma^2_{y^b}}\right)^{-1}, \tag{2.2}$$

using the variance of the model equivalents $\sigma^2_{y^b}$ and the observation error $\sigma^2_{y^o}$ for our scalar observation y^o.

Step B.

Step B. adjusts the ensemble spread by correcting the deviations of the updated ensemble members y_i^a from their mean $\overline{y^a}$:

$$y_i^a - \overline{y^a} = \alpha(y_i^b - \overline{y^b}) \qquad \text{with } \alpha = \left(\frac{\sigma^2_{y^o}}{\sigma^2_{y^o} + \sigma^2_{y^b}}\right)^{\frac{1}{2}} \tag{2.3}$$

The deviations of the updated ensemble members from their mean, $y_i^a - \overline{y^a}$, are proportional to the deviations in the prior ensemble, scaled by the proportionality factor α based on the variances of background and observation. α is always less than one, which means a reduction in the spread and, therefore, uncertainty.

Step C.

The computed state and spread adjustment are now projected to the model space by computing each member's observation space increment $\delta y_i = y_i^a - y_i^b$. This increment in

---

[1] NCAR - National Center for Atmospheric Research

observation space is now projected to the state space with a linear regression using the covariances between the ensemble members. This also allows for an update of unobserved variables based on the covariance between variables for each state variable j within the state vector x. The increment for each $x_{i,j}$ is given by:

$$\delta x_{i,j} = \frac{cov(x_j^b, y^b)}{\sigma_{y^b}^2} \cdot \delta y_i \tag{2.4}$$

$\delta x_{i,j}$ can be added to the state to yield $\overline{x^a}$, but it only represents the additional information from assimilating a single observation. Assuming uncorrelated observation errors, one can add each impact to the state sequentially and use the updated ensemble to assimilate the next observation.

2.3.2 Localization and inflation

Covariance localization

As shown in Equation 2.4, the EAKF accuracy strongly depends on the prior covariance matrix. However, the covariances can be suboptimal as a result of small ensemble sizes, model errors, non-linearities and various other factors (Anderson, 2012).

Covariance localization

Localization of the background error covariance matrix is applied to mitigate spurious correlations and reduce sampling errors caused by the limited size of the ensemble. The ensemble matrix is a low-rank approximation, as the affordable number of ensemble members is much lower than the degrees of freedom in the model (Evensen et al., 2022). As dynamical effects can propagate in space but are primarily local, correlations closer to the observation are less affected by sampling errors than long-distance correlations. Therefore, today's primary approach to covariance localization is distance-based. It is assumed that there is also no sampling error with zero covariances. Therefore, a correlation function is chosen that supports covariances locally and filters remote and, thus, noisy covariances. A Schur product[2] of the selected function and the background error covariances calculated from the ensemble is taken (Houtekamer & Mitchell, 2001) to ensure that the product function is again a covariance function (Gaspari & Cohn, 1999). A commonly used correlation function is the one by Gaspari and Cohn (1999), also known as the Gaspari-Cohn-function. This function reduces the covariances by a factor smaller than 1, decreasing with distance. The factor reaches zero after a certain length scale that can be set according to the expected influence of each observation type. To ensure a smooth spatial update, the covariances of neighbouring observations should overlap. For visible reflectance assimilations, Kugler et al. (2023) assumed a cutoff distance of 20 km with an observation density of 10 km, while Scheck et al. (2020) set the cutoff to 35 km with a superobbed observation every 12 km.

---

[2] Schur product - an elementwise matrix multiplication

Background error inflation

Another way of accounting for sampling errors and addressing unrepresented error sources is background error inflation. It is a direct way to increase the variances prior to or posterior to the assimilation step. Prior inflation is applied to mitigate variance underestimation caused by model errors, while posterior inflation methods mitigate sampling errors caused during the assimilation. Inflation algorithms can generally be multiplicative or additive (Gharamti, 2018). This study uses a posterior multiplicative inflation method called relaxation to prior spread (RTPS), as described by Whitaker and Hamill (2012). The ensemble standard deviation of the posterior is relaxed back to the prior:

$$\sigma_a \leftarrow (1-\alpha)\sigma_a + \alpha\sigma_b \tag{2.5}$$

where σ_a is the spread (defined as standard deviation) before and σ_b after the assimilation. α is a changeable parameter defined between 0 and 1, where 1 indicates a relaxation to 100 % of the prior spread.

Sample Error Correction - SEC

Inflation and covariance localization are the most common methods for dealing with deficiencies of the sampled covariances. Furthermore, Anderson (2012) described a statistical model called Sample Error Correction (SEC). Prior information about the distribution of the correlation between a state and an observed variable is used to reduce the error of the analysis. This method applies a correction factor between 0 and 1 to all state variables during the assimilation. An offline Monte-Carlo simulation precomputed the factor for different ensemble sizes and sample correlations. As the SEC factor is applied similarly to a traditional localization, it effectively reduces the sensitivity to the localization strength (Anderson, 2012).

2.3.3 The role of the observation error

Another essential role in the EAKF plays the assigned observation error σ_o and the corresponding variance σ_o^2. Depending on the accuracy of the observation indicated by this parameter, the observation is given more or less weight during the assimilation. The assigned observation error should contain the instrument and representativity error and account for errors related to the observation operator (Janjić et al., 2018). Furthermore, the error estimate can be inflated and used to account for neglecting correlated errors and other deficiencies of the DA system. Therefore, the observation error value for the optimal update is not clear apriori. Visible reflection and cloud fraction are bounded quantities, and ideally, the observation error decreases towards the bounds to avoid including impossible values in the assumed Gaussian distribution (Scheck et al., 2020). Finding a dynamic estimate is not straightforward, and most studies apply a constant observation error estimate for simplicity (Kugler et al., 2023; Scheck et al., 2020; Schröttle et al., 2020).

2.4 Verification methods

The verification scores used in this thesis are the mean absolute error (MAE) and the root mean squared error (RMSE) of the ensemble mean, as well as the probabilistic fraction skill score (pFSS).

Mean absolute error (MAE) and root mean squared error (RMSE)

To compute spread-error-relationships, the root mean squared error (RMSE) for one observation type was used:

$$RMSE = \left\langle \left(\frac{1}{N} \sum_{i=1}^{n} F_i - O \right)^2 \right\rangle, \tag{2.6}$$

where F is the forecasted field for each member i, O the observed field and $\langle . \rangle$ indicates the average over all observations. For verifying vertical profiles as was done by Kugler et al. (2023), the mean absolute error, MAE, was computed:

$$MAE_z = \left\langle \frac{1}{N} \sum_{i=1}^{n} |F_{i,z} - O_{i,z}| \right\rangle, \tag{2.7}$$

with MAE_z being the value of the MAE for a specific model level z. MAE_z is the mean over both horizontal dimensions indicated by $\langle . \rangle$, and it is also a mean over the ensemble members N.

2.4.1 Fraction Skill Score - FSS

Grid point verification, like the RMSE, can introduce double-penalty errors when small displacements of events in the forecast lead to it being penalized twice (Necker et al., 2023). This is especially problematic for high-resolution forecasts, as the intrinsically limited small-scale predictability for precipitation or clouds leads to many displacement errors (Bachmann et al., 2020). Thus, to avoid unfair forecast assessment, applying object-based spatial verification methods, using some sort of scale separation or considering a neighbourhood, is recommended (Necker et al., 2023).

Deterministic FSS

The Fraction Skill Score (FSS), as introduced by Roberts and Lean (2008), is a method to verify deterministic forecasts while alleviating the influence of double penalty errors by comparing neighbourhood fractions (Necker et al., 2023). It is based in neighbourhood probabilities NP_j for each grid point J:

$$NP_j = \frac{1}{M} \sum_{m=1}^{M} BP_{j,m} \qquad \text{with } BP_j = \begin{cases} 1, & \text{if } F_j \geq t \\ 0, & \text{if } F_j < t \end{cases}, \tag{2.8}$$

with M being the number of grid points in the neighbourhood of a grid point j and BP_j being the binary probability for a threshold t - similarly defined as the cloud fraction in 3.1. The Fraction Brier Score FBS compares forecasted fields f to observed ones o using neighbourhood fractions over the number of all grid points to verify J:

$$FBS = \frac{1}{J} \sum_{j=1}^{J} (NP_{j,f} - NP_{j,o})^2 \tag{2.9}$$

The worst FBS, $WFBS$, corresponds to no overlap between the observation and the forecast:

$$WFBS = \frac{1}{J} \left[\sum_{j=1}^{J} NP_{j,f}^2 + \sum_{j=1}^{J} NP_{j,o}^2 \right] \tag{2.10}$$

The deterministic Fraction Skill Score FSS_{det} is then the fraction of $WFBS$ and FBS and allows measuring the skill of a forecast for a certain neighbourhood or window size (Necker et al., 2023):

$$FSS_{det} = 1 - \frac{FBS}{WFBS} \tag{2.11}$$

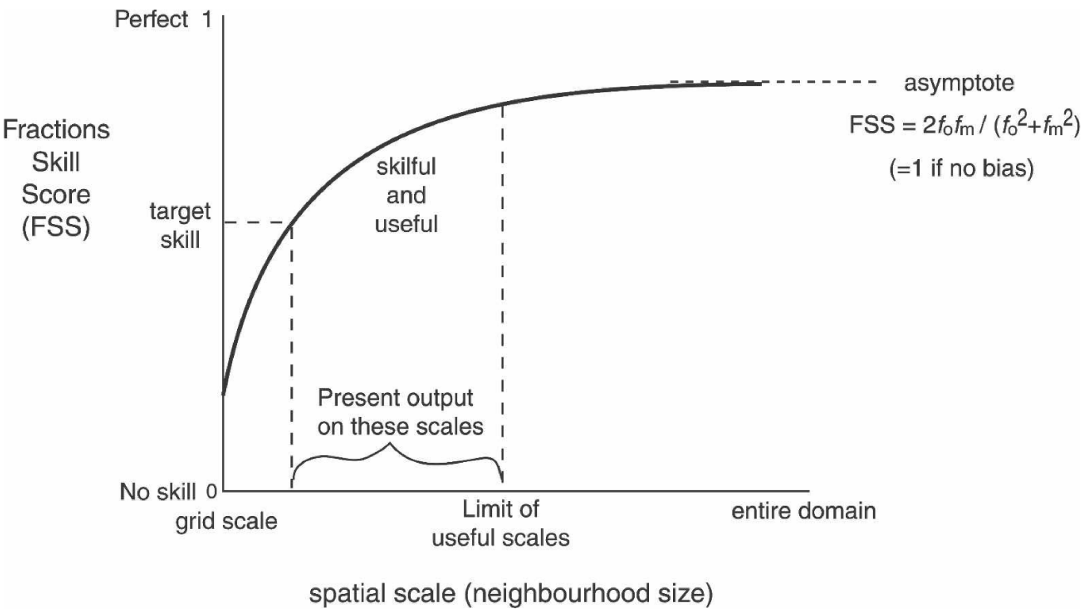


Figure 2.3: Schematic illustration of the FSS: The y-axis is the FSS value ranging from 0 to 1, and the window size is on the x-axis. The unbiased FSS approaches asymptotic 1 with increasing spatial scale. Figure taken from Roberts and Lean (2008).

As shown in figure 2.3, the FSS has a range of 0, being the worst, to 1, the perfect match between forecast and observations. As the size of the squares to compute the fractions increases, the FSS will approach an asymptotic value depending on the ratio of observed and forecasted events (Roberts & Lean, 2008). This so-called asymptotic FSS (AFSS) is based on the FB_{det}, the deterministic frequency bias:

$$AFSS_{det} = \frac{2 \cdot FB_{det}}{1 + FB_{det}^2} \qquad \text{with } FB_{det} = \frac{fo}{fm}, \tag{2.12}$$

where f_o and fm are the fractions of gridpoints exceeding the threshold t in the domain and the total number of gridpoints J for the observation and the forecast, respectively. Therefore, a frequency bias smaller than one indicates that the event occurs more often in the forecast than it was observed and vice versa if it is greater than 1. The asymptotic FSS is less sensitive to biases from small frequencies than the frequency bias and indicates a bias if it is smaller than its perfect value 1 (Roberts & Lean, 2008).

Probabilistic FSS - pFSS

There is more than one way to account for the ensemble dimension when computing a probabilistic FSS, with most having advantages and disadvantages. Necker et al. (2023) tested four approaches, and the $pFSS$, based on combining neighbourhood and ensemble probabilities for the forecast, is used in this study. It is also implemented in an open-source software package by Wolfgruber (2024). It utilizes Neighbourhood Ensemble Probabilities $NEP_{j,f}$:

$$NEP_{j,f} = \frac{1}{N} \sum_{n=1}^{N} NP_{j,n,f} = \frac{1}{M} \sum_{m=1}^{M} EP_{j,m,f}, \tag{2.13}$$

where N is the number of ensemble members and EPj, m, f is the ensemble probability, with M again being the number of grid points in the neighbourhood. To compute $pFSS$, $NP_{j,f}$ in equation 2.9 and 2.10 has to be replaced by $NEP_{j,f}$ from 2.13 (Necker et al., 2023):

$$pFSS = 1 - \frac{\sum_{j=1}^{J} (NEP_{j,f} - NP_{j,o})}{\sum_{j=1}^{J} NEP_{j,f}^2 + \sum_{j=1}^{J} NP_{j,o}^2}. \tag{2.14}$$

The AFSS for a probabilistic forecast, $pAFSS$, also has no commonly agreed upon definition (Necker et al., 2023). For this thesis, 2.12 was adapted with the ensemble mean of the frequency bias pFB:

$$pAFSS = \frac{2 \cdot pFB}{1 + pFB^2} \qquad \text{with} \quad FB_{pFB} = \frac{1}{n} \sum_{i=1}^{n} \frac{fo}{fm_i} \tag{2.15}$$

Uncertainty of the pFSS

As far as the author knows, there is no common, agreed-upon, and straightforward way to express the pFSS's uncertainty. Different analysis and forecast times have been aggregated to alleviate issues from sampling errors due to a small number of observation-forecast pairs. Depending on the applied threshold and the verified field, the pFSS might suffer from a limited sample size. Especially for large-scale cloud fractions, only a few observations cover the whole domain. Time can be added as another dimension to make the computation more robust. The different valid times T are aggregated, as has been done with the ensemble members in equations 2.13 and 2.14, yielding

$$pFSS_{time} = 1 - \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}(NEP_{j,t,f} - NP_{j,t,o})}{\sum_{t=1}^{T}\sum_{j=1}^{J} NEP_{j,t,f}^{2} + \sum_{t=1}^{T}\sum_{j=1}^{J} NP_{j,t,o}^{2}}. \tag{2.16}$$

3 Experimental design

The following chapter describes the experimental design applied in this study. It gives a detailed overview of how cloud fraction is computed and assimilated (section 3.2). The last section of the chapter, section 3.4, focuses on the empirically obtained observation error estimate.

3.1 OSSE

An Observing System Simulation Experiment (OSSE) is designed to use (new) data assimilation algorithms and proposed data inputs to investigate their potential impact. OSSEs are conducted in an entirely simulated environment: instead of using actual observations, OSSEs draw their observations from a simulated atmosphere with simulated errors. Therefore, the actual state of the virtual atmosphere is known precisely, unlike reality, and can be used to verify OSSE experiments (Errico & Privé, 2018).

OSSE workflow

An OSSE typically comprises a **nature run**, simulating the 'Truth', and is used to create **simulated observations** with an observation operator. Those are then assimilated using a **data assimilation algorithm** and appropriate error modelling to create an analysis. The analysis is then advanced in time with a NWP **forecast model** (Masutani et al., 2010).

To alleviate issues with systematic errors between modelled and observed clouds, a perfect model OSSE is used in this study, applying an identical model setup for creating the nature run and the forecasts. As the lack of model error can lead to a minor observation impact, this setup only shows a relative influence of the proposed ideas. Figure 3.1 depicts the workflow with components mentioned above linked to the models and algorithms used in this thesis, following closely the experimental design by Kugler et al. (2023).

3.1.1 WRF

The NWP model used in this study is the Weather Research and Forecast (WRF) model, version 4.3, as described by Skamarock et al. (2021). It was set up with the ARW (Advanced Research WRF) solver, a mass-conserving, fully compressible, Eulerian, non-hydrostatic equations solver. It features a terrain-following hybrid sigma-pressure vertical coordinate and uses horizontal Arakawa C-grid staggering. This grid staggering type separates the evaluation of velocities and masses. East-west velocity components are evaluated at the centres of the left and right grid faces, and south-north components at

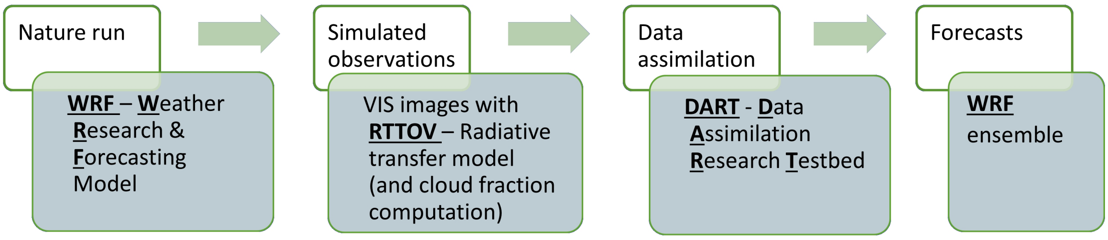


Figure 3.1: The OSSE workflow for a data assimilation experiment using DART and WRF.

the centres of the upper and lower grid faces. To integrate forward in time, it uses a time-split 3rd-order Runge-Kutta scheme and different time steps for acoustic and gravity modes (Skamarock et al., 2021). WRF was run in an idealized setup for the ensemble forecasting and the nature run, with identical settings.

The domain has a 2 km resolution and comprises 256 × 256 mass gridpoints. The model top is set to 21.5 km with 101 vertical levels on fixed hybrid eta-coordinate levels. The whole domain features homogenous, flat terrain and double periodic boundary conditions. To simulate solar radiation, a summer day in early July was chosen, and the domain is centred around latitude 47°N and longitude 7°E. The used parameterizations include Morrison et al.'s 2-moment scheme for microphysics, 5-layer thermal diffusion for the land-surface model, the Mellor-Yamada-Nakanishi-Niino (MYNN) schemes for the boundary and surface layer, and the RRTMG option for the atmospheric long- and shortwave radiation.

Nature run

The nature run was initialized using a modified sounding from Schröttle et al. (2020), shown in Figure 3.2, following Kugler et al. (2023). A high amount of CAPE (2670 J/kg) and low amount of CIN (26 J/kg) allow the triggering of convection with relatively small perturbations. To create randomly scattered convective cells, at the initialization at 7 UTC, the potential temperature at every grid point is perturbed with a 0.2 K standard deviation of a uniform distribution. This leads to a series of ordinary thunderstorms, as seen in Figure 3.3. The resulting convective cells have a random location, grow, interact dynamically and dissipate again. At 11 UTC, the around 20 resulting storms are in different development stages, leading to swiftly growing model errors and low predictability. This continues onward to around 15 UTC when convection starts to decay.

Forecast ensemble

Following Kugler et al. (2023) and Schröttle et al. (2020), the ensemble was created by applying two kinds of perturbations for each member. Large-scale errors are represented

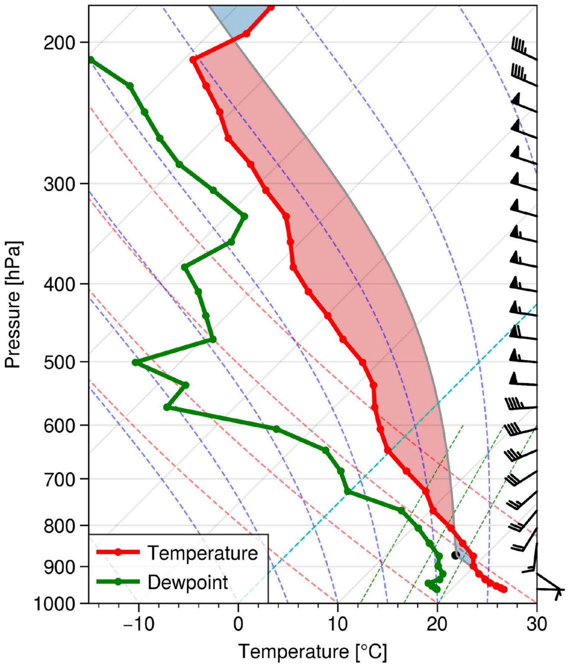


Figure 3.2: Skew T–log-p diagram depicting the initial conditions for the nature. Profiles of temperature, dewpoint, and a parcel-lifting curve are shown averaged over the domain. Figure taken from Kugler et al. (2023).

with vertically auto-correlated profile perturbations: the initial sounding's (Figure 3.2) temperature, moisture, and wind are perturbed. Every 20th vertical level is taken from the original profile and randomly perturbed with a 0.25 K standard deviation for temperature and 2 % of the initial value for relative humidity and wind. Vertically auto-correlated profiles are then created by interpolating between those random numbers. Random, small-scale noise was added to the vertical velocity and temperature fields at the lowest levels and exponentially decreased to zero with height. The 40-member ensemble exhibited a small uniform temperature spread at initialization time, which grew considerably over time, especially as the developing convection modified the profiles. The horizontally averaged spread (standard deviation of the ensemble) profile for temperature and water vapour mixing ratio are shown in Figure 3.4 for 12 UTC. The water vapour mixing ratio spread (Figure 3.4 b.) is reaching two peaks, one at the boundary layer and one slightly above 2.5 km height, while the peak spread in temperature (Figure 3.4 d.) can be found between 5 and 7.5 km with 0.8 K. As the ensemble is created, the convective indices of each ensemble, such as its level of free condensation or CAPE value, differ initially. With time, the ensemble spread is not evolving equally in different heights.

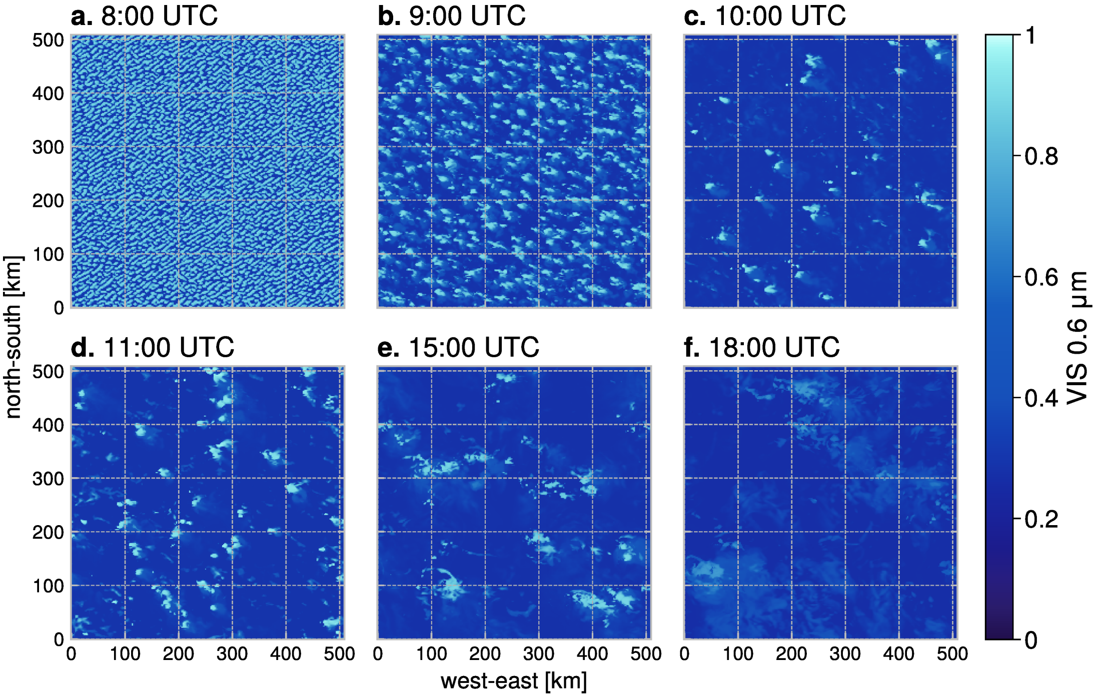


Figure 3.3: Visible 0.6-μm satellite images of nature run at different times.

3.1.2 DART

The Data Assimilation Research Testbed, DART, is an open-source community facility for ensemble DA, developed and maintained by NCAR[1]. It is possible to use DART with different NWP models and several implemented DA schemes, including the EAKF (Anderson et al., 2004). This study uses DART version 10.9.1, coupled with WRF.

The assimilation settings followed Kugler et al. (2023): Temperature, water-vapour mixing ratio, the dry air mass in the column, geopotential, three-dimensional wind components (U, V, W), and cloud water and ice mixing ratio were updated during assimilations. As ensembles tend to be underdispersive and become even more so when observations are assimilated, a type of posterior inflation was used: Relaxation to prior spread (RTPS) with $\alpha = 0.9$, meaning that posterior spread was inflated to 90 % of the prior. Covariance localization was applied to reduce spurious correlations and observation density. As all performed assimilation featured observations without height information, localization was only enforced horizontally with a Gaspari–Cohn function. To ensure a smooth spatial update, the overlap of the impact of neighbouring observations is necessary. Therefore, localization, as explained in section 3.2.1, was adapted depending on the

---

[1] NCAR - National Center for Atmospheric Research

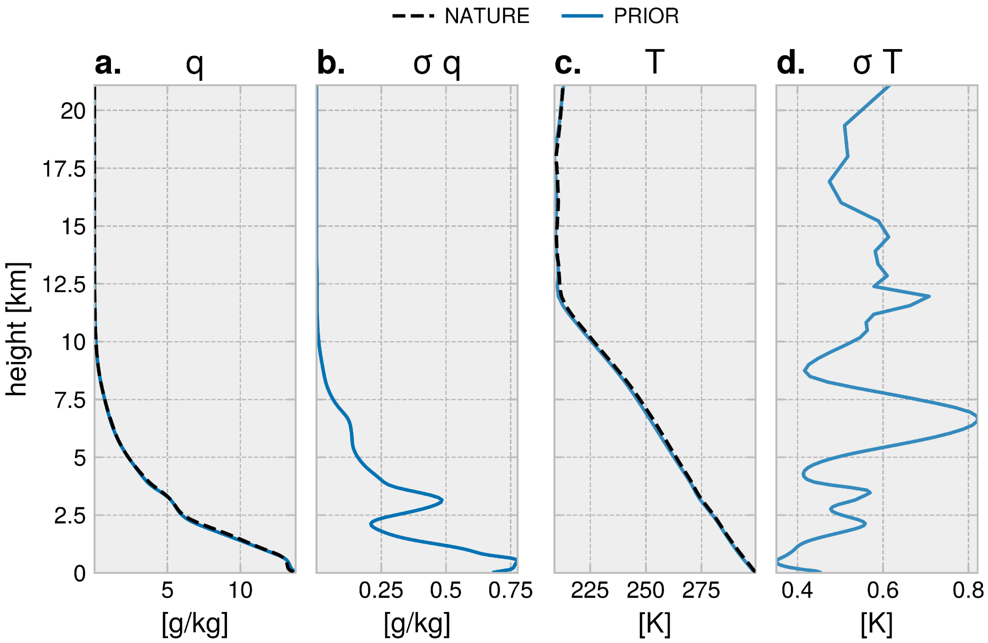


Figure 3.4: Horizontally averaged profiles over the whole domain for a. water vapour mixing ratio (q) and c. temperature (T) at 12 UTC. Blue is the ensemble mean of the free ensemble with 40 members, and the dashed black line is the nature run profile. Corresponding ensemble spread profiles are shown in panels b. and c., respectively.

assimilated scale. The setup also featured a sampling error correction (Anderson, 2012) to alleviate problems caused by a too-small ensemble size.

Creating synthetic images with RTTOV

The forward operator to create the synthetic observations in this study is the RTTOV-MFASIS (Saunders et al., 2018). The settings applied to RTTOV need to be consistent with the WRF setup, so it was centered around the same coordinates. Furthermore, it requires the satellite's zenith and azimuth angle, which were set to 45° and 180°, respectively. The surface elevation above sea level was set to 490 m.

Pre-computing the forward operator

Unlike direct VIS assimilation, cloud fraction assimilation is not implemented in the DART algorithm. This means that the synthetic observation must be created beforehand for the nature run and each ensemble member and then added to a so-called 'obseq'-file, which can otherwise be created directly by DART. It then includes the observed value drawn from the nature run, the value for each ensemble member, and coordinates of the observation, as well as the corresponding observation error variance. Each observed information is catalogued to a specific observation type. New types, such as cloud fraction, must be added before compiling DART, documented by Anderson et al. (2004) and Kershaw (2021).

3.2 Cloud fraction assimilation

3.2.1 Cloud fraction computation

With the RTTOV-MFASIS observation operator, a synthetic visible reflectance (0.6 μm) field from a WRF output file with the exact same resolution can be created. The corresponding cloud fraction values based on the visible field are easily computed: the two-dimensional field is divided into N squares depending on the area size over which the cloud fraction shall be computed. Each square is transposed to a binary field by a predefined threshold t, and the cloud fraction value C is then obtained by dividing the sum of pixels in the square exceeding the threshold $\sum p_i$ by the total amount of pixels in the square A:

$$C = \frac{\sum_{i=1}^{A} p_i}{A} \text{ with } p_i = \begin{cases} 1, & \text{if } p_i \geq t \\ 0, & \text{if } p_i < t \end{cases} \tag{3.1}$$

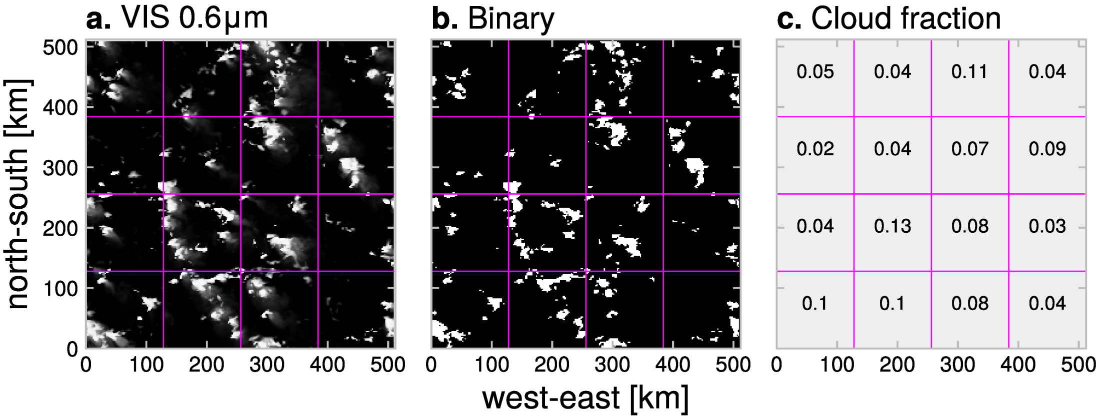


Figure 3.5: Illustration of Equation 3.1 with an example of 16 squares: a. is the visible reflectance field, b. the same field after setting all values below the threshold to 0 and c. shows the cloud fraction value for each square. Cloud fraction was computed with a visible reflectance threshold of 0.6.

Equation 3.1 is illustrated by Figure 3.5: a binary field from the visible reflectance image is created with a threshold and C computed for each sub-square. The cloud fraction variable is naturally bound between 0 and 1. It can be computed over differently sized areas, ranging from the whole domain to only a few pixels, which is depicted in Figure

3.6. Panel b.-d. show cloud fraction values calculated over 256x256 km, 128x128 km, 64x64 km, 32x32 km and 16x16 km, respectively. Smaller scales (8x8 km or 4x4 km) can be computed but are not effectively resolved by the model, and the cloud structure on those fine scales is unlikely to be realistic.

Cloud fraction threshold

Cloud fraction C is heavily dependent on its threshold value - it also distinguishes the cloud fraction computation from averaging over a subdomain or other forms of superobbing. The visible reflectance is a highly nonlinear function of the optical depth. Due to scattering effects, ice clouds have a generally lower visible reflectance than water clouds with the same water content (Scheck et al., 2020). Therefore, setting a high enough visible reflectance threshold allows isolating convective cells (values above the threshold) from cirrus clouds (reflectance below the threshold). As C gives no height information during the assimilation, excluding pixels with low visible reflectance seems beneficial. This thesis uses a threshold of 0.6 visible reflectance for all cloud fraction computations, as has been for verification purposes in Kugler et al. (2023) and is similar to the 0.5 threshold used in Scheck et al. (2020).

Covariance localization and cutoff

To minimize spurious correlations, an appropriate covariance localization needs to be applied. The cutoff length scale depends on the spatial scale of the cloud fraction observation. This thesis uses a localization cutoff radius R of the same length as one side of the cloud fraction computation area square, schematically illustrated by Figure 3.7. This means the covariances will decrease following a Gaspari-Cohn function and fade to zero at two times the radius (dashed circle). This leads to an overlap of localizations from neighbouring observations, as indicated by the grey square in Figure 3.7 and ensures a smooth spatial update.

The localization cutoff is adjusted for each cloud fraction observation depending on its area of computation, leading to non-zero covariances at the domain boundaries. As DART is unaware of WRF's double periodic boundary conditions, it cuts off localization at the boundaries, which can cause numerical errors and suboptimal data assimilation updates. Only observations simulated from a smaller subdomain were used during the assimilation to avoid any issues. For the initial 512x512 km domain, the smaller assimilation zone was set to 256x256 km and is shown in Figure 3.6 as an orange rectangle.

Simulating observations

Cloud fractions can be computed directly from the nature run and assimilated as a perfect observation. Realistically, there is an instrument error with the visible reflectance field that is then not accounted for. To simulate observations from the nature run, each pixel of the computed visible reflectance field is perturbed with 3 % error, as has been done in Kugler et al. (2023). The cloud fraction is then calculated from the disturbed field.

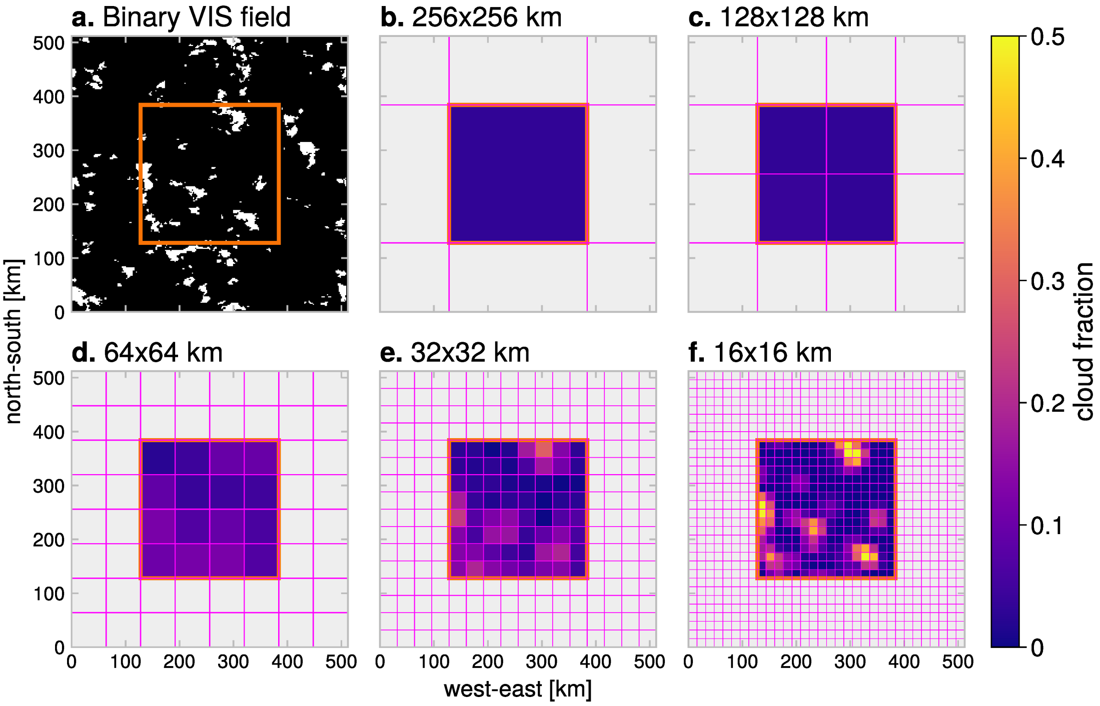


Figure 3.6: Cloud fraction computation on different scales: a. shows the visible field for reference, b.-d. cloud fraction values for this field calculated over differently sized areas indicated by the magenta grid. Only observations in a smaller inner domain (orange rectangle) are shown. Cloud fraction was computed with a visible reflectance threshold of 0.6.

3.2.2 Multiscale cloud fraction assimilation

As Figure 3.6 demonstrates, cloud fraction calculation is possible on multiple scales from one image. This opens up the possibility of updating multiple scales. As a sequential filter is applied, the assimilated information must be uncorrelated across scales to avoid suboptimal updates.

The difference approach

One way to do so is to assimilate differences between cloud fraction scales instead of cloud fraction values themselves, as illustrated by Figure 3.8. The first and coarsest scale, in this case, a 256x256 km area cloud fraction, is taken as is. The next finer scale, 128x128 km, is defined as the difference to the previous 256x256 km cloud fraction. Iteratively, differences can be computed for each scale down to 16x16 km cloud fraction and together add up again to the first coarse observation. This difference calculation is added to the cloud fraction computation function and applied to the ensemble and the nature run.

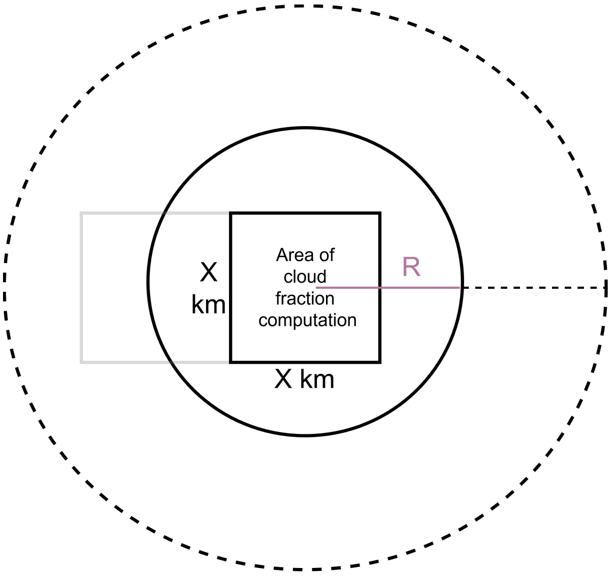


Figure 3.7: Schematic illustration of the covariance localization used for cloud fraction observations: for a cloud fraction computed over an area (X times X km), a cutoff radius R of X km is defined (pink line). The covariances are decreased following a Gaspari-Cohn-function before fading to zero at two times R, depicted by the dashed circle. The grey square indicates a neighbouring observation.

3.2.3 Cycling setup

The assimilation cycling was set up the same way for each conducted experiment and is schematically shown in Figure 3.9. The assimilation window was set between 12-13 UTC based on the convective developments depicted in Figure 3.3, with five assimilation cycles every 15 minutes. In an idealized experimental design, the start of the cycling should be set to a time when the ensemble has developed enough spread to ensure a consistent spread-error-relationship. As the ensemble spread is defined as the standard deviation of the ensemble, it should, in theory, be equal to the RMSE of the ensemble mean in Equation 2.6 (Eckel & Mass, 2005). To check the spread-error-relationship, the RMSE and the spread for the water vapour mixing ratio (q), visible reflectance (VIS), a coarse cloud fraction scale (256x256 km) and a fine-scale one (16x16 km) were computed for each timestep between 12 and 13 UTC and shown in Figure 3.10. The RMSE and spread of each observation type were divided by the corresponding maximum RMSE to allow a comparison of all fields in one figure. While humidity and VIS fulfill the criterion, the coarse cloud fraction type exhibits an RMSE that is too small for the corresponding ensemble spread and vice versa for the finer-scale cloud fraction. Especially the coarse scale cloud fraction spread and RMSE are not robust, as not a lot of observation-forecast pairs are available.

After the last assimilation at 13 UTC, a free forecast was performed for 5 hours to

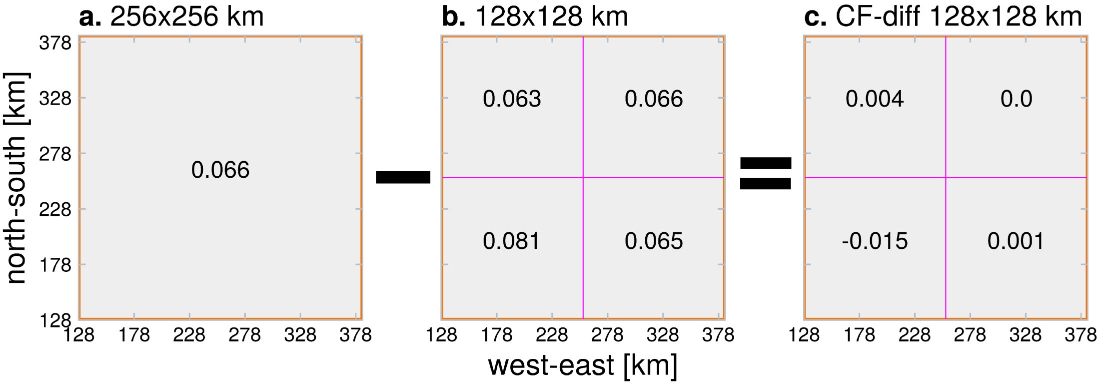


Figure 3.8: Cloud fraction differences: a. shows the coarsest scale (256x256 km) cloud fraction, b. the next coarsest 128x128 km and c. the cloud fraction difference between a. and b.; Areas of computation are indicated by the magenta grid. Only observations in a smaller inner domain (orange rectangle) are shown.

evaluate impact of the experiments with time.

3.3 Benchmark experiment - visible reflectance assimilation

The cloud fraction simulation OSSEs are directly compared to a benchmark experiment featuring direct visible reflection assimilation. The benchmark experiment follows Kugler et al. (2023) as closely as possible. Instead of using pre-computed observations as this study does with cloud fractions, DART was compiled with RTTOV, which initializes an observation type for visible reflectance observations before the DART compilation. The forward operator is called during the assimilation and the visible reflectance observations are created automatically. RTTOV-MFASIS was set up with the same settings as described in section 3.1.2.

Pixel-by-pixel assimilation of the VIS image is not recommended, as the distance between assimilated observations is too small to ensure uncorrelated observation errors. Like Kugler et al. (2023), the benchmark experiment in this thesis applies observation thinning and only considers one pixel every few kilometres.

3.4 Empirical observation error variance estimate

An empirical estimate was attempted as there is currently no estimate of the observation error variance for cloud fraction. Kugler et al. (2023) assigned an error of 3 % to the visible reflectance assimilation and for creating the synthetic observations. The same error estimate was used in this thesis to simulate the visible reflectance instrument error. To investigate how this error is transposed to an error in cloud fraction, the visible reflectance

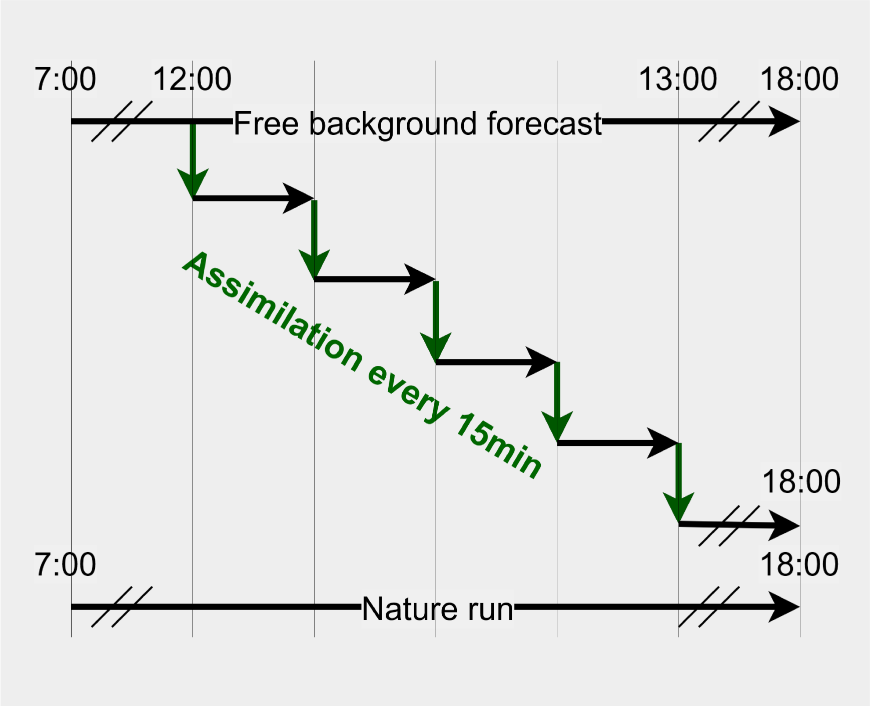


Figure 3.9: Schematic illustration of the assimilation cycling: black arrows indicate a free forecast and green ones refer to the assimilation. After the first assimilation at 12 UTC, the analysis is forecasted for 15 minutes and serves as the background for the next assimilation at 12:15 UTC. This continues until 13 UTC, resulting in five cycles.

image was disturbed 10^6 times with 3 %, and for each realization, the corresponding cloud fraction was calculated for each cloud fraction area computation. The standard deviation over all realizations is then taken as an estimate for the observation error. Clear sky observations that only occur on fine scales exhibited an extremely low observation error variance (0 up to 10^{-23}). If, for example, a 16x16 km cloud fraction observation was zero before the visible field's perturbation, the VIS field's added error is never enough to impact the cloud fraction computation due to the high threshold of 0.6. Almost 60 % of all cloud fraction observations are zero on the 16x16 km scale, leading to low observation error variance estimates as seen in Figure 3.11 panel a). To inflate the error on finer scales, those clear sky cloud fraction observations have been neglected, leading to the results depicted in Figure 3.11 panel b) and the mean estimates σ_o^2 given in Table 3.4. The model excluding clear sky observations has been used as a constant error model for each scale. Estimates depend greatly on scale, with σ_o^2 for the cloud fraction without clear sky observations exhibiting a roughly logarithmic dependence and attributing less error to coarser scale observations, as depicted in Figure 3.11 b).

However, the cloud fraction error variance estimate is not only scale-dependent but

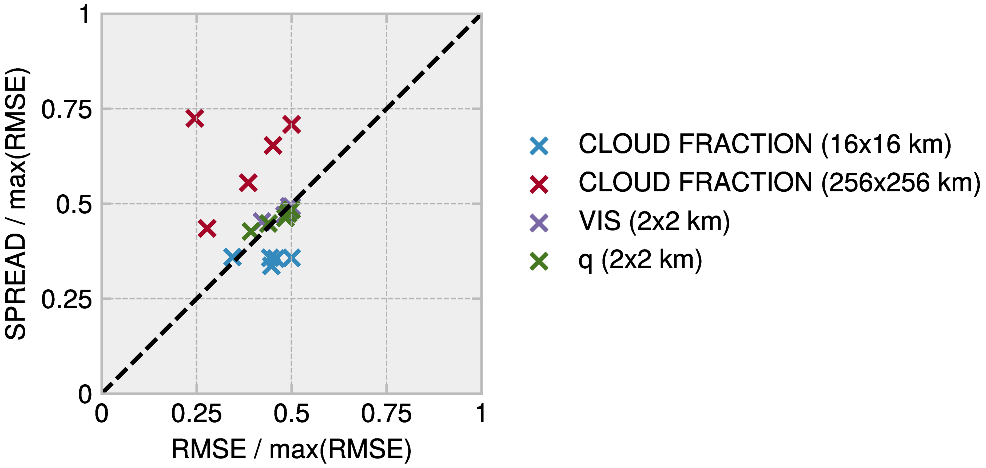


Figure 3.10: The spread-error-relationship is illustrated by showing the mean RMSE divided by the maximal RMSE for each observation type on the y-axis compared to the corresponding mean ensemble spread, again divided by the maximum RMSE for every observation type on the x-axis. The selected observation types are a coarse cloud fraction (256x256 km), a small-scale one (16x16 km), visible reflectance (VIS) and the vertical mean water vapour mixing ratio (q). Each cross indicates one 15min timestep between 12-13 UTC.

also depends on the cloud fraction value. This is shown in Figure 3.12 with bins of cloud fraction values. Bigger areas of computation, in this case, 256x256, 128x128 and 64x64 km, lead to cloud fraction values being always below 0.4 and have been interpolated from 0.4 to 1. For finer scales, the cloud fraction computation exhibits a lower error estimate for extreme values, which are thus less error-prone than values in mid-range. For a bounded quantity like cloud fraction, this is recommended to exclude impossible values from the distribution (Scheck et al., 2020).

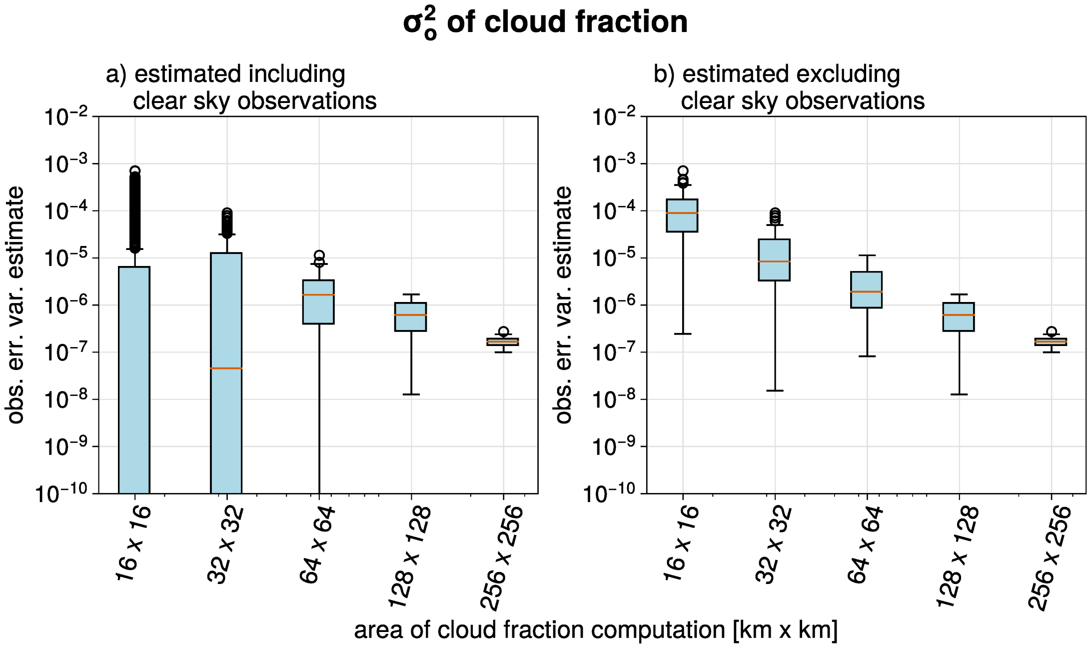


Figure 3.11: Observation error variance estimate for different scales of cloud fractions computed from disturbed nature visible reflectance images from 12-13 UTC. Panels a) and b) show the distribution as box plots for each scale if clear sky observations are included or excluded in the computation, respectively. The mean σ_o^2 is given in Table 3.4.

Scale	σ_o^2 including clear sky observations	... excluding them
256x256 km	$1.69 \cdot 10^{-7}$	$1.69 \cdot 10^{-7}$
128x128 km	$6.87 \cdot 10^{-7}$	$6.87 \cdot 10^{-7}$
64x64 km	$2.44 \cdot 10^{-6}$	$3.01 \cdot 10^{-6}$
32x32 km	$9.93 \cdot 10^{-6}$	$1.64 \cdot 10^{-5}$
16x16 km	$2.32 \cdot 10^{-8}$	$1.26 \cdot 10^{-4}$

Table 3.1: The mean observation error σ_o^2 estimated for different scales, for cloud fraction including and excluding clear sky observations.

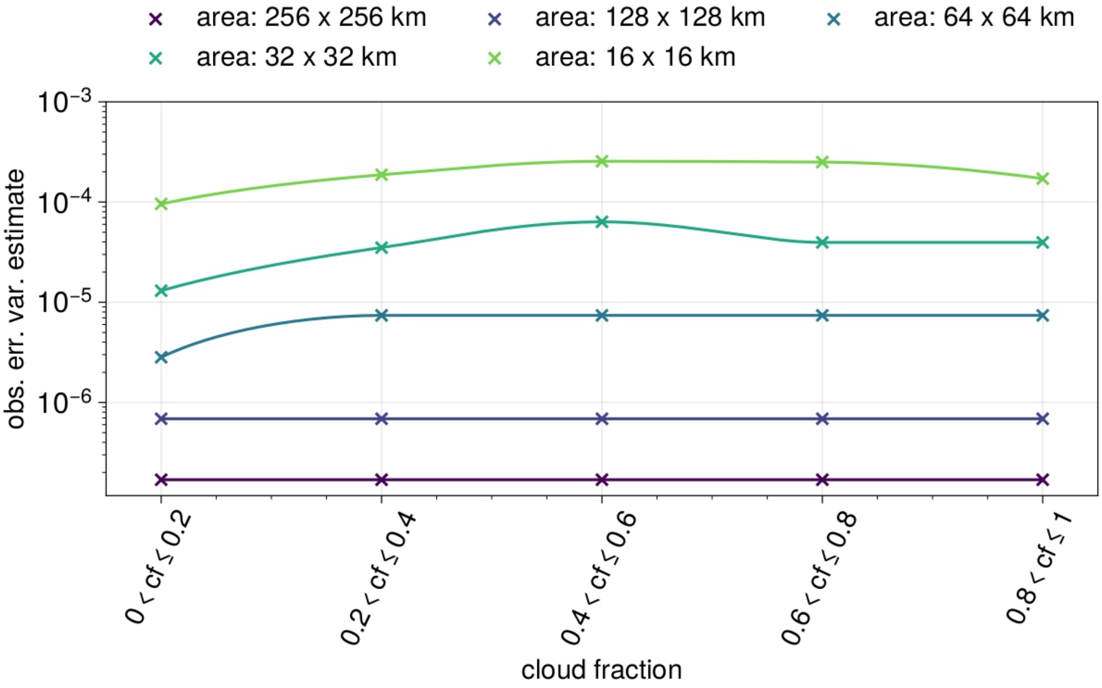


Figure 3.12: Observation error variance estimate as in Fig. 3.11, but for different bins of cloud fraction values and interpolated over the whole range from 0-1. Zero cloud fraction observations are neglected.

4 Results

This chapter provides results regarding single-scale cloud fraction assimilation experiments (section 4.1) as well as experiments featuring multiple scales (section 4.2). Potentially introduced biases by the multiscale approach are investigated in section 4.2.1, and an enhanced error model is proposed in section 4.2.2.

4.1 Single scale experiments

Name	Description	Error model	Assimilated cloud fraction scales	Number of assimilated observations per cycle
VIS_{16}	Thinned visible reflectance observations assimilated every 16 km with a localization cutoff of 32 km	constant, $\sigma_o^2 = 9.0 \cdot 10^{-4}$	/	256
CF_{con}	Single scale cloud fraction assimilation using a constant error variance estimate found empirically	constant (Fig. 3.11 b.), non-inflated	16x16 km	256
CF_{var}	Single scale cloud fraction assimilation using the empirical error variance estimate dependent on cloud fraction	varying (Fig. 3.12), non-inflated	16x16 km	256

Table 4.1: Overview over single scale experiments and the benchmark experiment.

Cycled experiments only featuring a single observation type on a single scale are listed in Table 4.1. VIS_{16} is the benchmark experiment, following Kugler et al. (2023), but using less dense observations to have the exact same observation count as two cloud fraction assimilation experiments: CF_{con} and CF_{var}, that have been conducted to assess the impact of cloud fraction assimilation on a single scale. CF_{con} is a cycled assimilation of 256 16x16 km cloud fraction observations with a constant observation error variance that was estimated excluding clear sky observations, as shown in Fig. 3.11 b) and Table 3.4. CF_{var} uses the cloud fraction value dependent error model based on Figure 3.12.

During the assimilation window from 12-13 UTC, the pFSS of the assimilated field, the 16x16 km cloud fraction, is improving for CF$_{con}$ and CF$_{var}$. As VIS and cloud fraction are closely related, the benchmark experiment also improves in this period. This is shown in Figure 4.1 a.-c., where the time evolution of the pFSS, computed with Equation 2.14 is shown. On all depicted windowsizes, 16, 32, and 64 km, both cloud fraction assimilation experiments exhibit improvements of around 0.05 pFSS during the free forecast period. For Figure 4.1's panels a.-c., the cloud fraction computation threshold was set to 0.4. The dependence of the pFSS computation on this threshold is shown in panels d.-f. It is based on the accumulated pFSS (Equation 2.16) computed for the assimilation window and the free forecast from 12-18 UTC. Improvements of around 0.05 in pFSS value to the free ensemble can be seen for all experiments, with the cloud fraction experiments slightly outperforming VIS$_{16}$, especially on higher cloud fraction thresholds. However, there is no significant dependence of the pFSS value on the cloud fraction computation threshold.

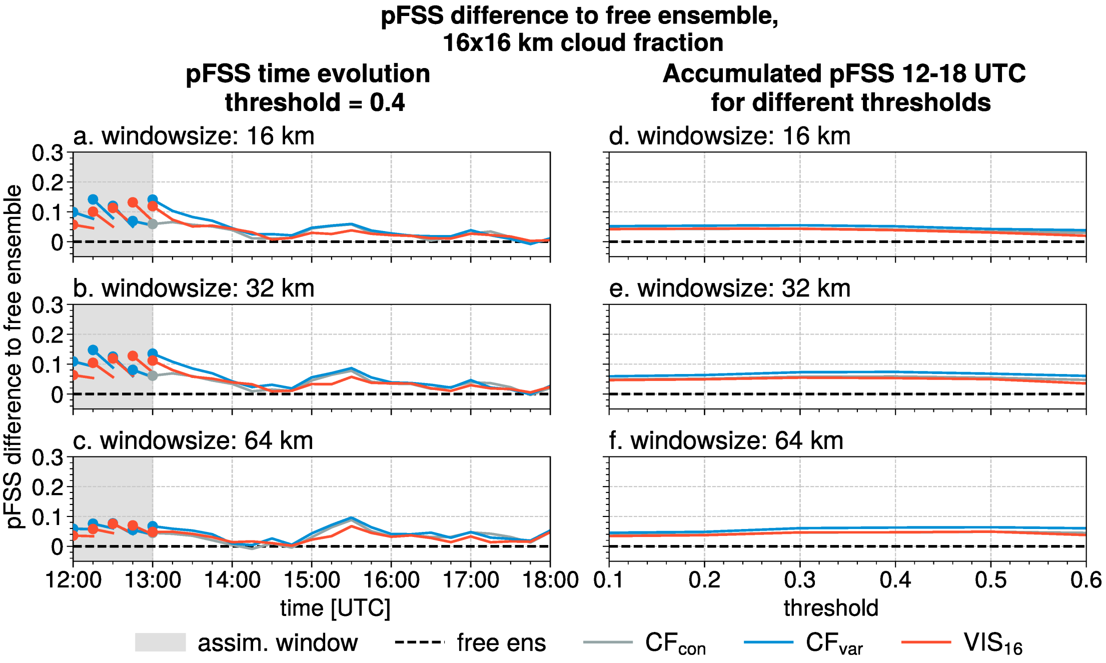


Figure 4.1: Difference in pFSS to the free ensemble (dashed black line) for 16x16 km cloud fraction. All experiments from Table 4.1 are shown. Panels a.-c. show the time evolution of the pFSS difference computed with a fixed threshold of 0.4 for three different windowsizes. Panels d.-f. depict the accumulated pFSS difference over the assimilation window and free forecast period (12-18 UTC) for three different windowsizes. The pFSS computation thresholds vary from 0.1 to 0.6 (shown on the x-axis).

Figure 4.2 shows the accumulated cloud fraction pFSS improvement to the free ensemble for a 16x16 km observation for a fixed threshold of 0.4. During the assimilation window, Figure 4.2 a., the benchmark experiment and both cloud fraction experiments exhibit improvements of over 100 % on the assimilated scale. The positive impact declines with increasing windowsizes for the assimilation window and during the free forecast period depicted in Figure 4.2 b. The overall improvement of accumulated pFSS for the free forecast period starting from 13:15 and ending at 18 UTC, where no longer observations are assimilated, is, therefore, smaller than for the assimilation window. CF_{var} performs best during the free forecast, exceeding the benchmark's improvement and CF_{con} by around 10 %. In neighbourhoods ranging from 16 to 160 km, CF_{var} is exhibiting the most skill. CF_{con} and also VIS_{16} are decreasing the pFSS by 1-2 % above 100 km scales. This might be purely coincidental, as the pFSS accumulated over the free forecast is heavily dependent on the last analysis. A robust estimate of the performance after the assimilation has stopped would need to be computed over many free forecasts starting at different analysis times. Another uncertainty factor is the limited sample size of the cloud fraction computation itself because only 256 observation-forecast pairs are available per timestep for this observation type.

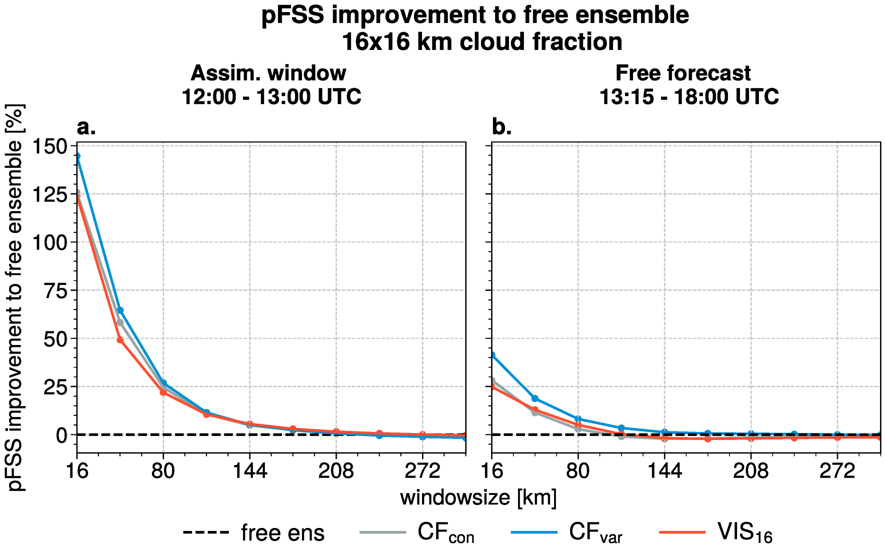


Figure 4.2: Accumulated pFSS difference to the free ensemble in percentage over a. the assimilation window (12-13 UTC) and b. the free forecast period (13:15-18 UTC). The pFSS was computed for cloud fraction (16x16 km area of computation) and a pFSS computation threshold of 0.4. The windowsize varies from 16 to 304 km (x-axis). All experiments from Table 4.1 are shown.

A more robust verification field is the visible reflectance field, as it allows for a bigger sample size. Figure 4.3 shows that it exhibits similar behaviour as the 16x16 km cloud fraction. The improvement of pFSS for VIS compared to the free ensemble is depicted over the assimilation window and the free forecast period in panels a. and b., respectively. It was computed for a fixed threshold of 0.6, which is also the threshold for the cloud fraction computation. CF$_{con}$ and VIS$_{16}$ perform similarly, most improving on small scales and exhibiting no or a slight negative impact from 130 km windowsize onwards. CF$_{var}$ outperforms the benchmark experiment for 2-km-neighbourhoods by around 2 %.

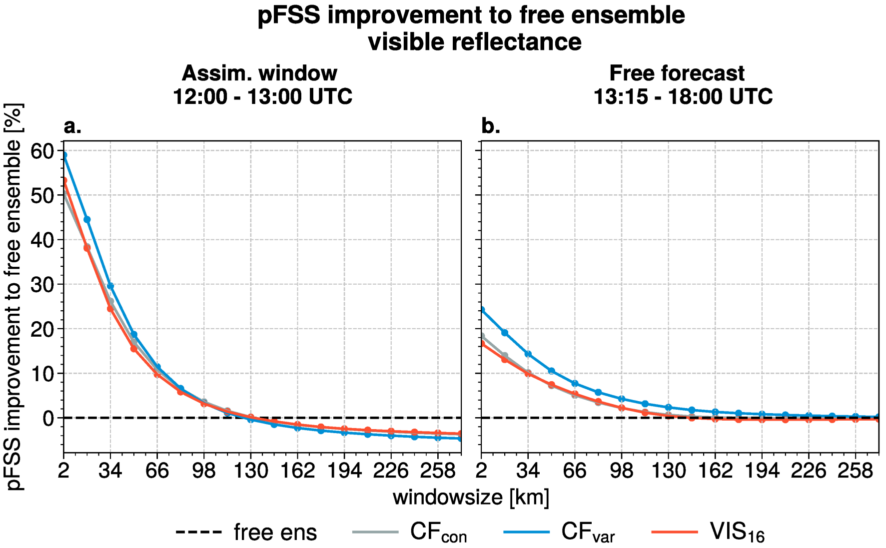


Figure 4.3: Accumulated pFSS difference to the free ensemble in percentage over a. the assimilation window (12-13 UTC) and b. the free forecast period (13:15-18 UTC). The pFSS was computed for visible reflectance and a pFSS computation threshold of 0.6. The windowsize varies from 2 to 274 km (x-axis). All experiments from Table 4.1 are shown.

A time series of the pFSS metric for the visible reflectance field gives Figure 4.4. Panels a.-c. depict the time evolution of the pFSS difference to the free ensemble for three windowsizes. Solid lines refer to the forecasts and dots to the analysis, indicating a clear improvement in pFSS with each assimilation cycle for the benchmark experiment VIS$_{16}$. The cloud fraction experiments are less consistent in their improvement in skill: the analysis of 12:45 UTC for CF$_{var}$ is only slightly exhibiting a positive effect, and the last analysis for CF$_{con}$ at 13 UTC is decreasing the skill compared to the free ensemble. The lead in performance seen in Figure 4.3 b. for CF$_{var}$ could be due to a very impactful 13 UTC analysis, where the difference between the two error models is most significant. The development of convective systems increases cloud fraction overall, and more mid-range values (around 0.5) are assimilated. All experiments can improve the forecast for up to 5 h lead time compared to the free ensemble.

Assimilation of cloud fraction can also benefit precipitation, as shown in Figure 4.4 panels d.-f. The 1-h accumulated precipitation pFSS, compared to the free ensemble, is depicted for three windowsizes and a fixed threshold of 2 mm. Due to the computation of a 1-h-accumulation, the full impact of the assimilation and corresponding improvement in skill is only seen after one hour. All experiments exhibit similar behaviour and have a positive effect of up to 5 h of lead time. On coarse scales, the impact is not strictly

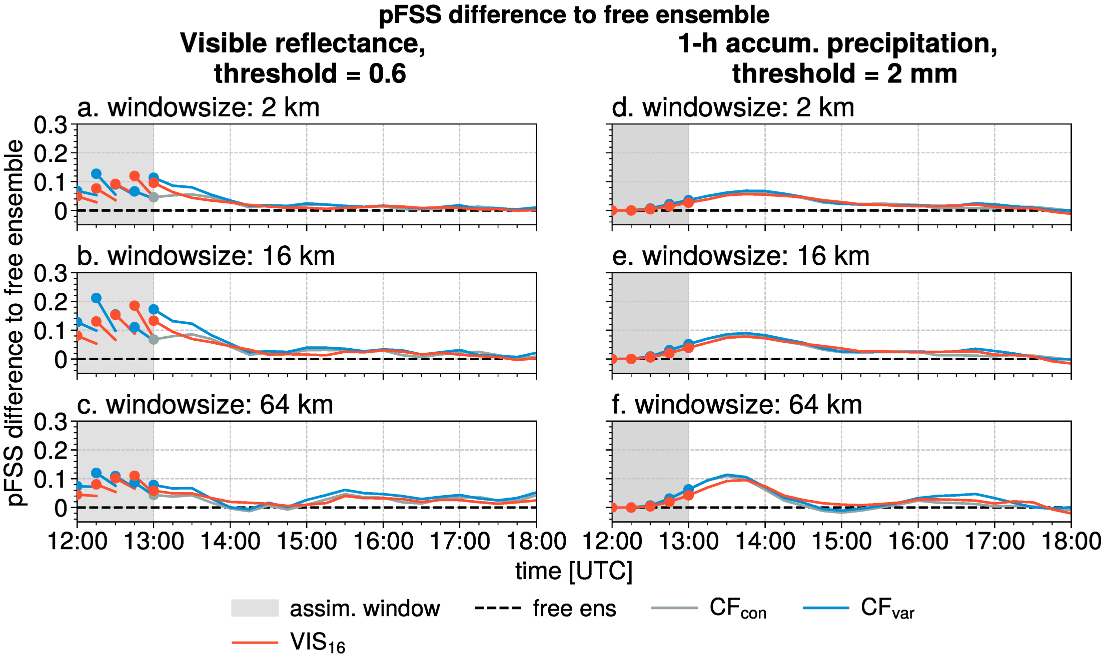


Figure 4.4: pFSS difference to the free ensemble (dashed, black line) for three different windowsizes. Panels a.-c. are computed for visible reflectance using a threshold of 0.6 while panels d.-f. refer to the 1-h accumulated precipitation with a threshold of 2 mm. All experiments from 4.1 are shown during the assimilation window (grey shading) and a free forecast with a 5 h lead time. Dots refer to an analysis, and solid lines to forecasts.

positive, with slight decreases in skill for the cloud fraction assimilation experiments around 2 h lead time.

4.2 Multiscale experiments

Four cycled experiments with assimilation of cloud fraction on more than one scale have been conducted and are listed in Table 4.2: MCF$_{diff1}$, MCF$_{diff2}$, MCF$_{seq1}$ and MCF$_{seq2}$. Adding multiple scales to the assimilation opens up another degree of freedom and a choice of which scales to assimilate at which frequency has to be made. MCF$_{diff1}$ and MCF$_{diff2}$ assimilate all available scales at each cycle to gain a better understanding of how this structural information impacts the analysis and forecast. MCF$_{seq1}$ only assimilates each scale once during the cycling, to explore the possibility of a more gradual approach by allowing 15-min-forecasts between each scale and therefore also adjustments by the model. MCF$_{seq2}$ is updating the state twice with all scales but only assimilates fine-scale cloud fractions in between. It is a combination of a more gradual update like MCF$_{seq1}$ while still utilizing the full potential of the cloud fraction computation similar to MCF$_{diff1}$. An overview of which scales are assimilated at which cycle for each multiscale experiment is given in Figure 4.5.

Name	Description	Error model	Assimilated cloud fraction scales	Number of assimilated observations per cycle
MCF$_{diff1}$	Multiscale cloud fraction assimilation using a constant error variance estimate; different scales are assimilated at the same assimilation time using the difference approach	constant (Fig. 3.11 b.), non-inflated	256x256 km, 128x128 km, 64x64 km, 32x32 km, 16x16 km;	341
MCF$_{diff2}$	As MCF$_{diff1}$, but with inflated observation error variance	constant (Fig. 3.11 b.), inflated by 100%	256x256 km, 128x128 km, 64x64 km, 32x32 km, 16x16 km;	341
MCF$_{seq1}$	Multiscale cloud fraction assimilation, using cloud fraction dependent error estimate and assimilating only one scale per assimilation cycle	varying (Fig. 3.12)	varying (Fig. 4.5)	varying between 1 and 256
MCF$_{seq2}$	Multiscale cloud fraction assimilation using a constant error variance estimate inflated by 100 %; at 12 and 13 UTC, all scales are assimilated with the differences approach and only the finest two in between.	constant (Fig. 3.11 b.), inflated by 100%	varying (Fig. 4.5)	varying between 320 and 341

Table 4.2: Overview over conducted multiscale experiments.

In MCF$_{diff1}$, each scale of cloud fraction observation was assimilated at the same timestamp using a constant error variance model without inflation. It also distinguishes it from MCF$_{diff2}$, where the observation error σ_o was inflated by 100 %. Both MCF$_{diff}$ experiments assimilate differences of cloud fraction as explained in section 3.2.2. MCF$_{seq1}$ assimilates only one scale per cycle, starting with 256x256 km cloud fraction computation area at 12 UTC, continuing with 128x128 km at 12:15 UTC, 64x64 km at 12:30 UTC, 32x32 km at 12:45 UTC and concluding with 16x16 km at 13 UTC. As no differences were assimilated, MC$_{seq1}$ is applying the cloud fraction value dependent error model,

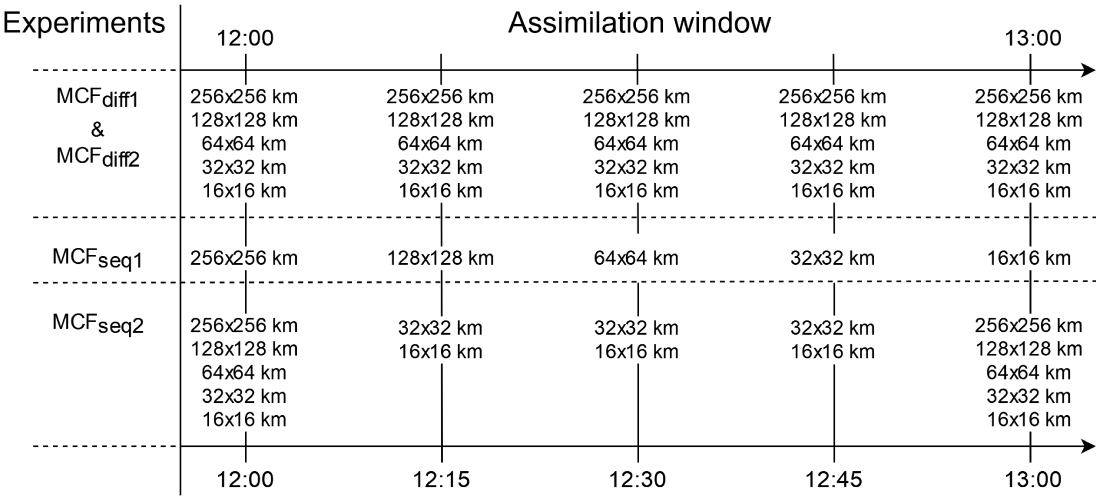


Figure 4.5: A schematic cycling flowchart following the structure as shown in Figure 3.9 for all multiscale experiments from Table 4.2. If more than one cloud fraction scale is listed, the difference approach from section 3.2.2 is applied.

non-inflated. The experiment MC$_{seq2}$ features a different sequential update: at 12 UTC, all scales from 256x256 to 16x16 km are assimilated using the difference approach, and from 12:15 to 12:45 UTC, only 32x32 and its 16x16 km difference are assimilated, before concluding the cycling with again assimilating all scales at 13 UTC. MCF$_{seq2}$ also applies the constant cloud fraction error model inflated by 100 % as was done with MCF$_{diff2}$.

Figure 4.6 panels a.-c. show a time series of the pFSS for visible reflectance of all multiscale experiments. All experiments offer similar improvements in skill during the 5-h-forecast compared to the free ensemble. The impact of assimilations in MCF$_{seq1}$ is negligible during the first cycles. The initial assimilation of very coarse scale cloud fraction is not expected to improve fine-scale analysis. MCF$_{diff1}$ improves the skill in general but seems to have difficulties on bigger windowsizes: MCF$_{diff1}$'s 13 UTC analysis is even reducing skill. In contrast, MCF$_{seq2}$ consistently improves around 0.1 pFSS on a 64-km-windowsize throughout the free forecast period. Like the single-scale experiments, the improvements during the assimilation cycling differ from analysis to analysis. All conducted experiments struggle with the 13 UTC analysis on bigger scales, similar to the behaviour seen for single-scale experiments in Fig. 4.4 c. MCF$_{seq2}$ is also not increasing skill on that particular analysis time but at least does not exhibit a negative impact like MCF$_{diff1}$ and MCF$_{diff2}$.

Regarding the 1-h-accumulated precipitation, the multiscale experiments show an overall positive effect on pFSS, as depicted in Figure 4.6 panels d.-f. MCF$_{diff1}$ and MCF$_{diff2}$ are exhibiting similar behaviour to VIS$_{16}$ in increasing skill directly after the assimilation window and decreasing the pFSS compared to the free ensemble after 1 to 2 h lead time. MCF$_{seq1}$ is assimilating the least observations, and its cycling only

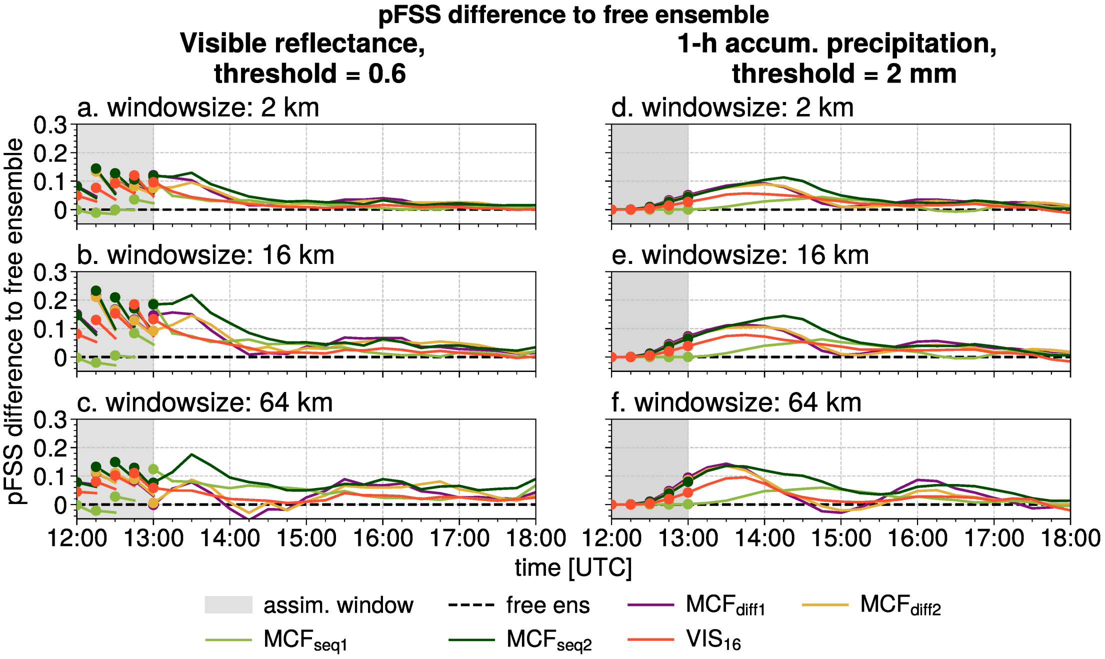


Figure 4.6: pFSS difference to the free ensemble (dashed, black line) for three different windowsizes: panels a.-c. are computed for visible reflectance, using a threshold of 0.6 and panels d.-f. refer to the 1-hour accumulated precipitation for a threshold of 2 mm. All experiments from Table 4.2 and the benchmark experiment VIS_{16} are shown during the assimilation window (grey shading) and a free forecast with 5-h lead time. Dots refer to an analysis, and solid lines to forecasts.

takes effect after 2 h lead time, consistently increasing the pFSS up to 0.05. The most consistent increase in skill is exhibited by MCF_{seq2}, with improved precipitation pFSS of around 0.1 for the entire shown free forecast time and all windowsizes. Especially on coarse scales, where all other experiments exhibit a slight decrease in skill during the free forecast period, MCF_{seq1} and MCF_{seq2} have a consistently positive impact.

Figure 4.7 showcases the accumulated VIS pFSS of the multiscale experiments for the assimilation window in panel a. and the free forecast period in panel b. For the latter, MCF_{seq1} and MCF_{seq2} are consistently improving the skill for VIS over all scales, with especially MCF_{diff2} also exhibiting the most improvement of 40 % on a 2 km neighbourhood. MCF_{diff1} and MCF_{diff2} perform similarly and improve pFSS on small scales by over 30 % compared to the free ensemble. However, they exhibit a negative impact on scales above 130 km.

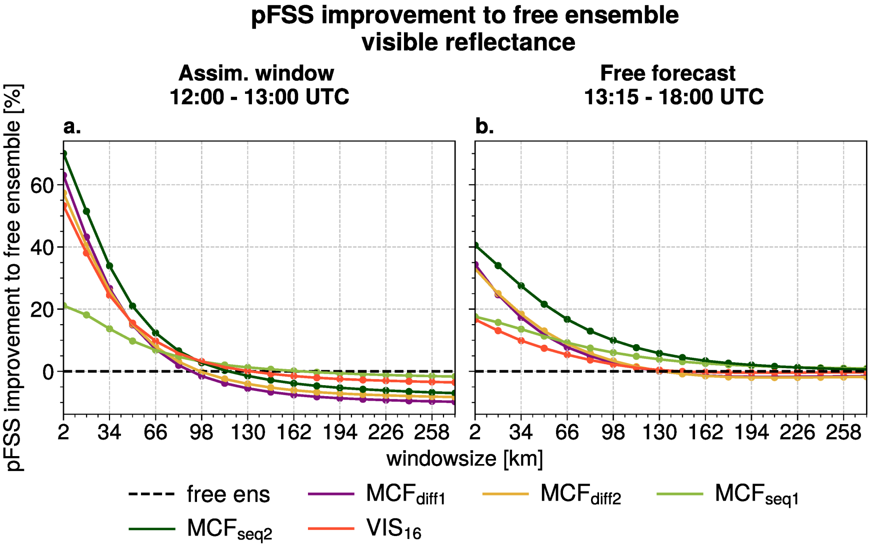


Figure 4.7: Accumulated pFSS difference to the free ensemble in percentage over a. the assimilation window (12-13 UTC) and b. the free forecast period (13:15-18 UTC). The pFSS was computed for visible reflectance and a pFSS computation threshold of 0.6. The windowsize varies from 16 to 274 km (x-axis). All experiments from Table 4.2 plus the benchmark experiment VIS_{16} are shown.

4.2.1 Spurious convection and a possible bias

Figure 4.8 shows the pFSS of the first and last analysis for all neighbourhoods. A definite trend can be seen during the cycling: especially MCF_{diff1}, MCF_{diff2} and MCF_{seq2} are adding clouds to the state with each cycle leading to a frequency bias below 1, meaning the model is exhibiting more clouds than the nature run. After a 2 h lead time, most incorrectly added clouds seem to have disappeared, as the different experiments fare similarly regarding frequency bias, which is illustrated in Figure 4.9. Panel a., the AFSS, is less sensitive to outliers and gives a general idea of how well an experiment approaches the unbiased value of 1. MCF_{seq1} is performing best here, as it introduces the least bias on large scales, while MCF_{seq2} is exhibiting a similar patterns as both MCF_{diff} experiments.

This can also be seen in the horizontally averaged mean absolute error of the water mixing ratio in Figure 4.10 c. and d. With each assimilation up to the depicted last analysis at 13 UTC, humidity was added to the vertical profile without any height information. Consequently, MAE in the horizontally averaged profile was not reduced but enhanced in certain areas. Especially, MCF_{diff1} and MCF_{diff2} add twice as much MAE as the other experiments at around 2.5 km height. Figure 4.10 panels a. and b. show that those two experiments also alter the temperature profile the most. The inflation of the observation error does not seem to affect the profile modification a lot, as it only slightly dampens it.

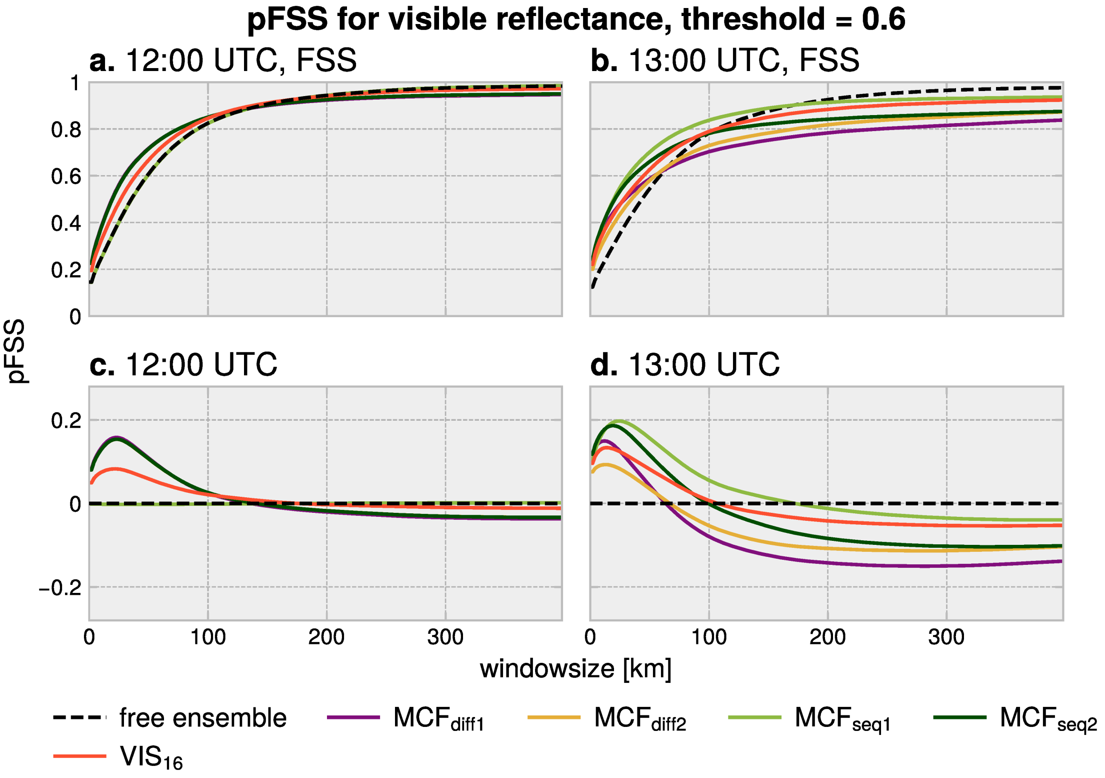


Figure 4.8: pFSS for visible reflectance (0.6 threshold) for two analysis times. Panel a. and b. show the pFSS for the first analysis at 12 UTC and the last analysis at 13 UTC, respectively, while c. and d. depict the same with the difference in pFSS to the free ensemble to enhance differences between the experiments. All experiments from 4.2 are shown, plus the benchmark experiment VIS_{16}.

Figure 4.11 depicts how the cloud fraction assimilation experiments alter the visible reflectance compared to the benchmark. It shows the probability field of the ensemble for the binary event of visible reflectance exceeding 0.6 μm. For all cloud fraction experiments, the peaks of the ensemble probability are less precisely tracing the green outlines of the nature run than VIS_{16}. For example, the convective cell at around x = 240 km and y = 250 km was not captured accurately by MCF_{diff1} and MCF_{diff2}, while the benchmark experiment is exhibiting a peak of ensemble probability in that region. However, the smaller scale cells around x = 250 km and y = 350 km show little to no change in comparison to the free ensemble for VIS_{16}, in contrast to MCF_{diff1} and MCF_{seq2}. Like the other cloud fraction assimilation experiments, they add small-scale probability peaks at seemingly random locations. This behaviour is visually more pronounced with experiments updating coarse scale cloud fractions frequently.

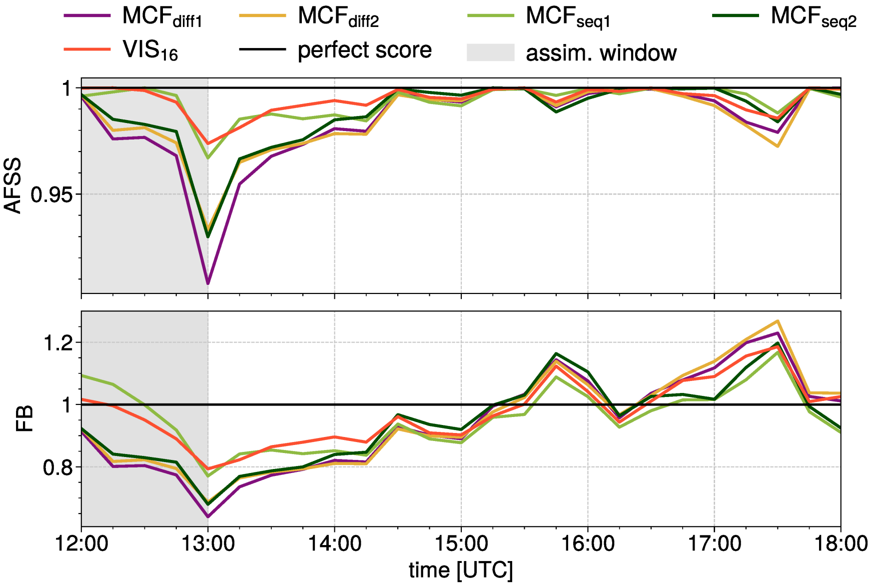


Figure 4.9: Looking into the assimilation bias: panel a. depicts the mean AFSS over all ensemble members for the first analysis at 12 UTC to the last analysis at 13 UTC (grey shading) and the free forecast until 18 UTC. Panel b. illustrates the mean frequency bias (FB) in the same way. The threshold for the computation is set to 0.6. All experiments from Table 4.2 and the benchmark experiment VIS$_{16}$ are shown.

4.2.2 An empirical error model for cloud fraction differences

As the empirical error model estimate, described in Figure 3.11 is based on cloud fractions, it might provide a suboptimal error estimate for assimilating cloud fraction differences. Therefore, another error model has been computed, shown in Figure 4.12. Clear sky observations have again been neglected, and the coarsest scale, 256x256 km, is taken as is because it is never assimilated as a difference. The new model exhibits a different behaviour: the scale dependence is not roughly logarithmic but attributing more error to 128x128 km and 32x32 km. Altogether, the mean estimated σ_o^2 values, illustrated in Figure 4.12, are slightly higher than for cloud fractions as seen in Figure 3.11 b). One multiscale experiment, MCF$_{seq3}$, has been computed to explore the impact of the better-suited error estimate.

MCF$_{seq3}$, as described in Table 4.2.2, is identical with MCF$_{seq2}$, except for the used observation error estimate. The same scales listed for MCF$_{seq2}$ in Figure 4.5 have been assimilated. A comparison of the pFSS time evolution between both experiments is shown in Figure 4.13.

Figure 4.13's a. and b. show the time evolution of the pFSS values for visible reflectance for MCF$_{seq3}$ compared to MCF$_{seq2}$ and the benchmark experiment. MCF$_{seq3}$ exhibits only slight changes, positively and negatively, to MCF$_{seq2}$. For example, the 12:45 UTC

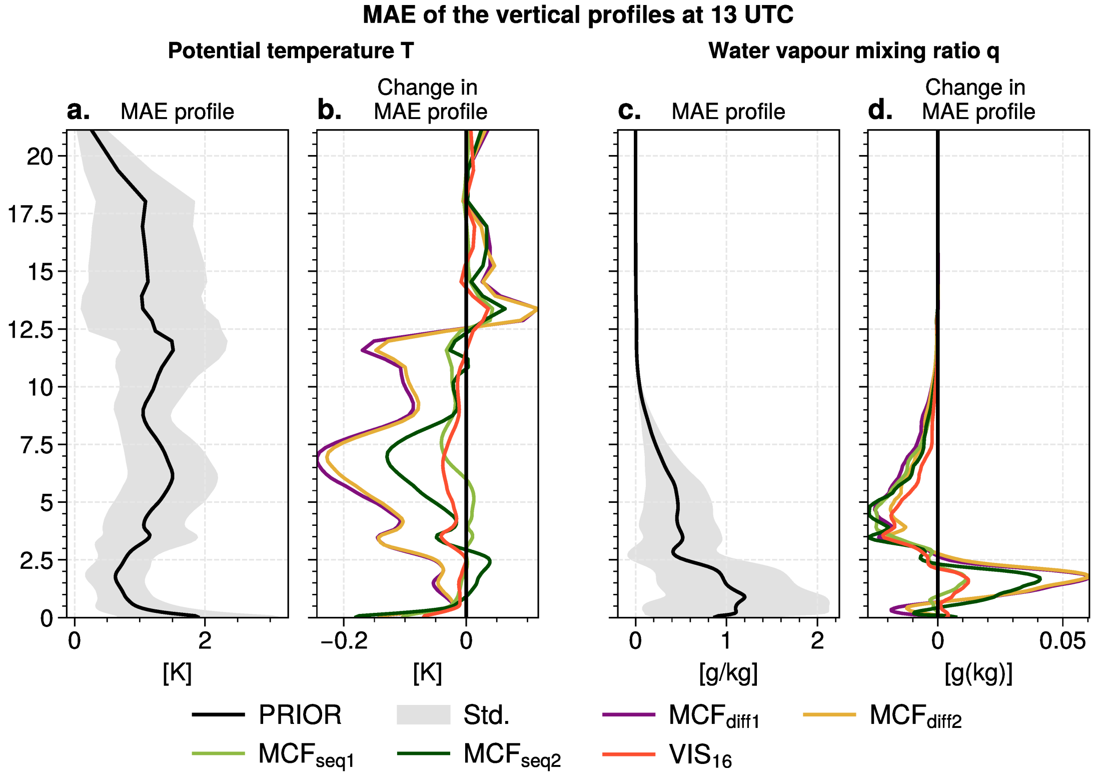


Figure 4.10: MAE of the horizontally averaged profiles; Panel a. and b. show the potential temperature MAE in K, c. and d. the water vapour mixing ratio q in g/kg. Panel a. and c. depict the MAE profile of the free ensemble at 13 UTC, with the grey shading indicating the standard deviation over all columns, while b. and d. show the change in MAE for all experiments from 4.2 plus VIS$_{16}$ at 13 UTC; negative values indicate an MAE reduction.

analysis of MCF$_{seq3}$ in Figure 4.13 a. did improve the pFSS value less than MCF$_{seq2}$, but for coarser scales, the 13 UTC analysis of MCF$_{seq3}$ is performing best. Even though MCF$_{seq3}$ does not overall perform better, it does not add as much bias during the assimilation period, as seen in Figure 4.13 c. The FB is significantly closer to the perfect score at 13 UTC for MCF$_{seq3}$. This can also visibly be seen in Figure 4.14, where the ensemble probability for VIS $> 0.6\ \mu m$ is shown. The amount of spurious convective events popping up is visibly reduced for MCF$_{seq3}$.

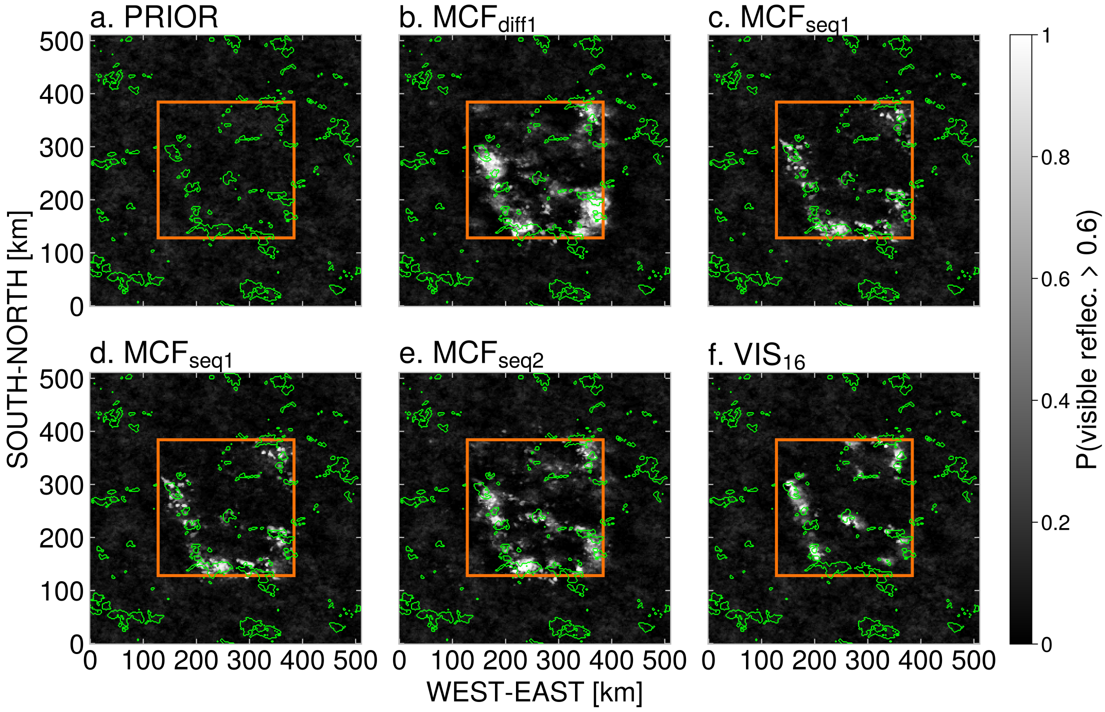
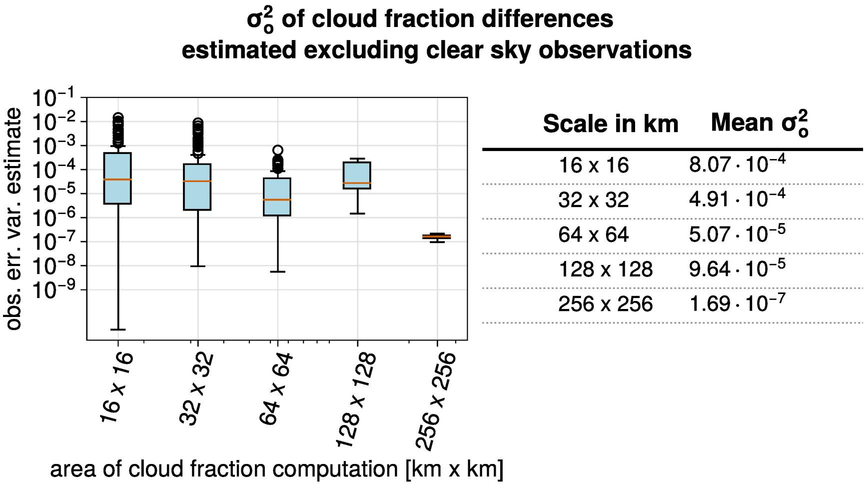


Figure 4.11: The ensemble probability of the visible reflectance field exceeding 0.6 at 13 UTC (last analysis). Panel a. illustrates the free ensemble, b.-e. depict cloud fraction experiments as listed in 4.2 and panel f. the benchmark experiment VIS$_{16}$. The orange box shows the inner domain, where observations have been assimilated. The green overlay is the outline of the visible reflectance binary field (threshold 0.6) of the nature run at the same time.



Scale in km	Mean σ_o^2
16 x 16	$8.07 \cdot 10^{-4}$
32 x 32	$4.91 \cdot 10^{-4}$
64 x 64	$5.07 \cdot 10^{-5}$
128 x 128	$9.64 \cdot 10^{-5}$
256 x 256	$1.69 \cdot 10^{-7}$

Figure 4.12: Observation error variance estimate σ_o^2 for cloud fraction differences for each scale; computed from disturbed nature visible reflectance images, neglecting clear sky observations, from 12-13 UTC. Box plots for each scale and the mean σ_o^2 values are shown.

Name	Description	Error model	Assimilated cloud fraction scales	Number of assimilated observations per cycle
MCF$_{seq3}$	Multiscale cloud fraction assimilation using a constant error variance estimate inflated by 100 %; at 12 and 13 UTC, all scales are assimilated with the differences approach and only the finest two in between.	constant (Fig. 4.12), inflated by 100 %	varying (Fig. 4.5, same as MCF$_{seq2}$)	varying between 320 and 341

Table 4.3: Introducing the multiscale experiment MCF$_{seq3}$.

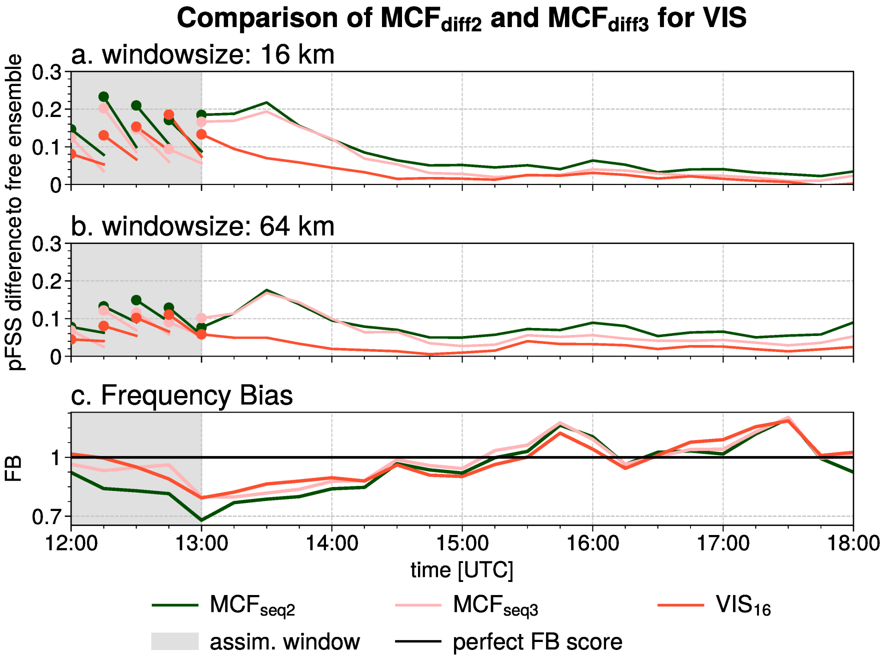


Figure 4.13: Visible reflectance pFSS evaluation of MCF$_{seq3}$ in comparison to MCF$_{seq2}$ and VIS$_{16}$; Panel a. and b. depict the pFSS difference to the free ensemble for a 16 and 64-km windowsize, respectively. Panel c. shows the time evolution of the frequency bias FB.

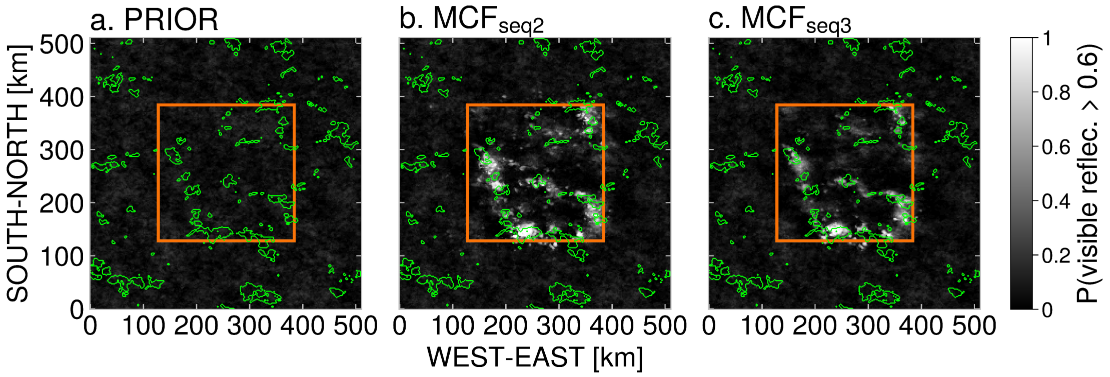


Figure 4.14: The ensemble probability of the visible reflectance field exceeding 0.6 at 13 UTC (last analysis). Panel a. illustrates the free ensemble, b. and c. depict MCF$_{seq2}$ and MCF$_{seq3}$, respectively. The orange box shows the inner domain, where observations have been assimilated. The green overlay is the outline of the visible reflectance binary field (threshold 0.6) of the nature run at the same time.

5 Discussion and conclusions

This study compares direct visible reflectance assimilation and assimilating structural information taken from VIS images: cloud fraction. Perfect-model OSSE experiments were conducted to evaluate the relative impact of cloud fraction assimilation, featuring an idealized, flat WRF domain with randomly distributed ordinary thunderstorms. Synthetic 0.6 μm visible reflectance images and cloud fraction values computed from those images on different scales have been assimilated with an Ensemble Adjustment Kalman Filter implementation. For the observation error, an empirical estimate was applied. Cycled experiments assimilating cloud fraction observations of a single scale and experiments exploring a multiscale approach were conducted. Cloud fraction differences have been assimilated in the multiscale experiments to ensure the independence of the observations and uncorrelated observation errors. A probabilistic Fraction Skill Score formulation, pFSS, was used to evaluate the results.

The main findings include:

- Single-scale cloud fraction assimilation exhibits similar or slightly improved skill compared to thinned VIS assimilation;
- Multiscale cloud fraction assimilations can outperform direct VIS assimilation in pFSS;
- Multiscale assimilations can introduce humidity biases and add spurious convection, but these issues are alleviated with a more appropriate error model;

Figure 5.1 summarizes the impact of multiscale cloud fraction assimilation. Panels a.-c. show how the assimilation affects the ensemble probability of each pixel exceeding 0.6 μm visible reflectance after the cycling. The benchmark experiment VIS_{16} enhances the probability similar to the structure of the green overlay, which indicates the binary event in the nature run. The cloud fraction assimilation seems to cause spurious convective cells throughout the domain, which can be seen in panels b. and c. However, a better-suited observation error model applied in MCF_{seq3} seems to lessen the effect and visibly enhance the ensemble probability. Figure 5.1 d. illustrates the impact on the accumulated pFSS over the free forecast for visible reflectance. Although MCF_{seq2} and MCF_{seq3} offer less accurate placements of individual convective cells, the forecast benefits from this uncertainty. Overall, the cloud fraction experiments outperform the benchmark on all scales. MCF_{seq3} is performing better than MCF_{seq2} on larger scales, which indicates a better representation of the overall convective activity in the domain.

Although the cloud fraction observation type is a bounded quantity and depends on a non-linear observation operator, its assimilation can substantially improve the skill of forecasting precipitation and clouds during the assimilation window and for up to 5 h free forecast lead time. Especially for multiscale experiments with different assimilation frequencies for different scales, cloud fraction seems a viable source of structural information leading to skill improvements across scales. It provides an alternative to thinned and direct visible reflectance assimilation, utilizing the observed information differently and subsequently impacting the overall convective activity in the domain.

The assimilation of a singular fine-scale cloud fraction in a cycled experiment (section 4.1) leads to improved skill for the assimilated variable and visible reflectance. Cloud fraction assimilation slightly outperforms visible reflectance assimilation compared to the benchmark experiment where the same number of observations was assimilated. The experiment utilizing an observation error estimate depending on the cloud fraction value, CF_{var}, performs best during the free forecast, exhibiting positive effects on coarse scales, where direct VIS assimilation has little to no positive impact. Compared to the free ensemble, cloud fraction assimilation consistently positively influenced precipitation forecasts.

As all conducted experiments share the same assimilation window, the free forecast results heavily depend on the last analysis. The 13 UTC analysis for the experiment featuring the constant error model, CF_{con}, was decreasing skill at the last analysis of the cycling, as seen in Fig. 4.4 b., leading to less pFSS improvement during the free forecast period. In this specific experimental design, the cloud fraction value-dependent error model proved useful, as it inflates the error of mid-range cloud fraction observations, where only a fragment of the computation area is covered in clouds and gives more weight to clear sky or full-cloudy ones. For the weather situation present at 13 UTC, many 16x16 km areas were only half covered with high VIS values, leading to many mid-range cloud fraction observations and, subsequently, a difference between the constant and the value-dependent error model. Clear sky observations were neglected in the computation of the error model, which might lead to a too-high error estimate on finer scales. Based on the computation and its threshold approach, clear sky observations would need to be assimilated theoretically with zero or almost zero as an error estimate.

Furthermore, errors are difficult to determine in an OSSE and deviate from the full real-world system. The proposed empirical error model accounts for the instrument error for simulated observations, and the representativeness error is irrelevant when assuming a perfect model and operator. However, it does not take uncorrelated random errors, systematic biases, or correlations between errors into account. Therefore, inflating the empirically obtained error might be beneficial for all values and not only for mid-range values. An inflation factor of 100 % for the constant error estimate has been assumed for the experiments MCF_{diff2}, MCF_{seq2} and MCF_{seq3}. As the optimal inflation value for the observation error is not clear apriori, further investigations and sensitivity studies would be needed to find an optimal estimate.

When assimilating all available scales in every cycling step, as done in MCF_{diff1} and MCF_{diff2}, the skill of visible reflectance on fine scales is increased by 20-30 % during the free forecast, as shown in Fig. 4.6 a.-c. However, a certain bias is added to the forecast, as pFSS values decline with increasing windowsize and indicate a negative impact of the assimilation on scales larger than 100 km. This multiscale approach also does not have a strictly positive effect on precipitation. Inflation of the observation error estimate helps to lessen the detrimental impact but does not change the pattern. Moisture is added across the vertical profile, and the horizontally averaged MAE of moisture increases with each assimilation. This indicates that assimilation is more efficient in adding clouds than removing them, resulting in a declining frequency bias during the assimilation window. After the last analysis, as the forecast model adjusts moisture and many created clouds disappear again, the frequency bias for all experiments is relatively similar after 1-2 h forecast time. The bias seen in this study might not be representative of different experimental designs - as the efficiency of cloud removal and creation is highly dependent on the amount of cloudiness already present in the domain.

Moreover, experiments have been conducted assimilating fewer scales of cloud fraction at the same time and updating the coarser scales only once or twice during the assimilation window, MCF$_{seq1}$ and MCF$_{seq2}$. As larger scales are less volatile with time than finer scales, they should also be assimilated less frequently. MCF$_{seq2}$ utilizes all scales only at the first and last analysis at 12 and 13 UTC, respectively. This doubles the pFSS improvement during the free forecast for visible reflectance and 1-h-accumulated precipitation compared to other multiscale experiments and outperforms the benchmark. A negative trend in the frequency bias during the cycling is still observed with MCF$_{seq2}$, but the horizontally averaged MAE for moisture is reduced by around 45 % compared to MCF$_{diff2}$. The experimental design can potentially be optimized further by changing at which time which scales are assimilated. Overall, the added structural information benefits the forecast in terms of pFSS if coarse scales are not assimilated too often. Hypothetically, this can be due to cloud fraction on those scales not changing rapidly and being stable over the short period of the assimilation window. Thus, updating too often leads to the introduced biases when combined with the assigned very small observation error. As the best-performing mix might depend heavily on the weather situation, determining the assimilated scales by the size and amount of observed convective cells with an automated algorithm or machine learning approach could potentially lead to great improvements in forecast skill.

Although the computed pFSS values might be similar for the benchmark and the cloud fraction assimilation experiments, the visually seen updates, as depicted in Figure 4.11 and 4.14, differ. As the experimental design contains, by default, many clear sky pixels in comparison to cloudy ones, the algorithm is more likely to add moisture and clouds in the whole updated area. The localization radius was adapted with scale, and the entire domain can be affected by the assimilation of one coarse scale cloud fraction observation. Spurious convection pops up depending on the initial fields of each ensemble member, leading to an ensemble probability field with no distinctive outline, contrary to the visible reflectance assimilation. This might not necessarily be a disadvantage, as small convective cells are intrinsically difficult to predict. The ensemble members exhibit different small-scale cloud patterns in regions where the nature run also has individual convective cells. Even as the exact location might not be perfectly met, this behaviour benefits the overall performance of the forecast as it captures the uncertainty in those regions. Especially with experiments utilizing less frequent coarse scale updates, the ensemble probability indicates areas with the possibility of small cloud structures occurring instead of not capturing them. This is potentially useful for probabilistic solar power production forecasting. However, too much added spurious convection, as seen with MCF$_{diff1}$, is degrading forecast skill again. As the observation error estimate is based on cloud fractions, it might be a suboptimal estimate for assimilating cloud fraction differences. An empirical error estimate based on cloud fraction differences is introduced in section 4.2.2 to find a more appropriate error model. A version of MCF$_{seq2}$ with this error estimate, called MCF$_{seq3}$, does improve pFSS values on larger neighbourhoods and visibly enhances the ensemble probability for visible reflectance at the last analysis. The amount of spurious convection is significantly reduced by the better-suited error model based on cloud fraction differences. However, the observation error estimates are highly dependent on their statistical sample, and averaging over various weather situations would be beneficial to achieve a more robust estimate, which was beyond the scope of this project. Furthermore, this study can only assess cloud fraction assimilation regarding relative impact, as there are no systematic representation differences between the nature and the forecast model in a perfect-model OSSE. If more error sources were considered, the benefit of having a higher ensemble spread in areas with small-scale convection could lead to significant forecast improvements.

The attempted proof-of-concept of assimilating structural information taken from a visible reflectance image was fulfilled, but further experiments and research are needed to test the robustness of cloud fraction assimilation. The experimental design of this study, with its randomly distributed convective cells, is intrinsically difficult to forecast, and the demonstrated positive impact of the assimilation in this idealized case indicates promising results for different weather scenarios. Ying et al. (2023) improved vortex positioning significantly using a multiscale assimilation approach based on wind information, and similar results might also be possible for cloud fraction assimilation in the presence of a synoptic-scale event. The experimental design of this study limits the coarser cloud fraction scales to a very small range, as every 256x256 km square is bound to contain at least some thunderstorms. A broader range of values might lead to different error model estimates but certainly to a different and possibly positive impact on the forecasts.

As visible reflectance and cloud fraction do not contain any height information, adding another observation type to the assimilation might prove beneficial in reducing the added error on the vertical profiles. The results shown in this project are generally very dependent on the creation and spread of the initial ensemble, which, in this case, is dominated by the differently timed onset of convective activity of the members and, therefore, their initial profiles. Assimilating soundings might improve the ensemble's vertical moisture distribution and facilitate the creation of clouds from cloud fraction assimilation at the correct height.

To summarize, assimilating cloud fraction on a single scale has a similar impact as direct visible reflectance assimilation. It enhances the visible reflectance and cloud fraction forecast on the assimilated scale and positively impacts precipitation. However, cycled experiments featuring multiple cloud fraction scales positively impact all assimilated scales, while direct VIS assimilation is improving skill mostly on fine scales. The multiscale approach has many degrees of freedom concerning the cloud fraction computation threshold itself, the error model, and the assimilation cycling and update frequency. Further studies are required to understand the behaviour of multiscale cloud fraction assimilation in different experimental designs and weather situations. However, cloud fraction is a viable alternative to direct assimilation and an easily implementable form of superobbing that has great potential for future applications.

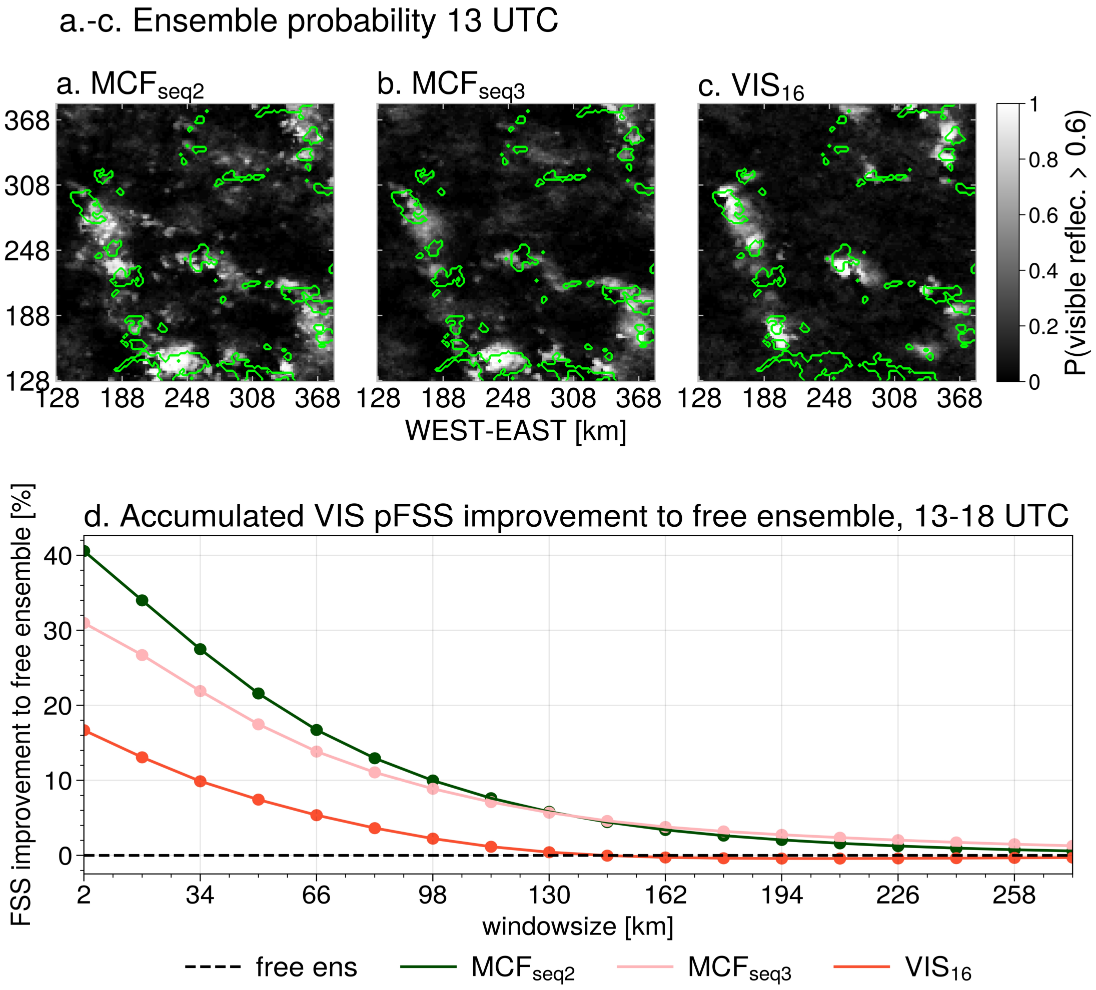


Figure 5.1: Summarizing the results of this thesis: Panel a.-c. show the ensemble probability of the last analysis (13 UTC) for VIS $> 0.6\ \mu m$ for the benchmark experiment VIS$_{16}$ and two multiscale cloud fraction experiments, MCF$_{seq2}$ and MCF$_{seq3}$. Only the inner domain, where observations are assimilated, is shown. The green overlay indicates VIS $> 0.6\ \mu m$ for the nature run. Panel d. is the accumulated pFSS improvement to the free ensemble for all neighbourhoods (x-axis) again for VIS$_{16}$, MCF$_{seq2}$ and MCF$_{seq3}$. The pFSS has been accumulated over the free forecast period (13:15-18 UTC) and calculated with a VIS threshold of 0.6 μm.

Bibliography

Abdalla, S., Isaksen, L., Janssen, P., & Wedi, N. (2013). Effective spectral resolution of ECMWF atmospheric forecast models.

Anderson, J. L. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903.

Anderson, J. L. (2012). Localization and sampling error correction in ensemble kalman filter data assimilation. *Monthly Weather Review*, *140*(7), 2359–2371.

Anderson, J. L., Hoar, T., Raeder, K., & Collins, N. (2004). *Data assimilation research testbed.* UCAR/NCAR - Computational; Information Systems Laboratory (CISL).

Bachmann, K., Keil, C., Craig, G. C., Weissmann, M., & Welzbacher, C. A. (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.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55.

Cintineo, R. M., Otkin, J. A., Jones, T. A., Koch, S., & Stensrud, D. J. (2016). Assimilation of synthetic GOES-r ABI infrared brightness temperatures and WSR-88d radar observations in a high-resolution OSSE. *Monthly Weather Review*, *144*(9), 3159–3180.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: A step-change in rainfall forecasting. *Meteorological Applications*, *23*(2), 165–181.

Duncan, D., Bormann, N., & Anderson, J. (2023). *Superobbing and finer thinning for all-sky humidity sounder assimilation* [ECMWF]. Retrieved March 27, 2024, from https://www.ecmwf.int/en/elibrary/81345-superobbing-and-finer-thinning-all-sky-humidity-sounder-assimilation

Eckel, F. A., & Mass, C. F. (2005). Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, *20*(3), 328–350.

Errico, R. M., & Privé, N. (2018). Some general and fundamental requirements for designing observing system simulation experiments (OSSEs).

Evensen, G., Vossepoel, F. C., & Van Leeuwen, P. J. (2022). *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem.* Springer International Publishing.

Eyre, J. R., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., & Pavelin, E. G. (2022). Assimilation of satellite data in numerical weather prediction. part II: Recent years. *Quarterly Journal of the Royal Meteorological Society*, *148*(743), 521–556.

Fabry, F., & Meunier, V. (2020). Why are radar data so difficult to assimilate skillfully? *Monthly Weather Review*, *148*(7), 2819–2836.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757.

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E. H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M., Köpken-Watts, C., & Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1191–1217.

Geiss, S., Scheck, L., de Lozar, A., & Weissmann, M. (2021). Understanding the model representation of clouds based on visible and infrared satellite observations. *Atmospheric Chemistry and Physics*, *21*(16), 12273–12290.

Gharamti, M. E. (2018). Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, *146*(2), 623–640.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., … Fujita, T. (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.

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., & Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278.

Jones, T. A., Otkin, J. A., Stensrud, D. J., & Knopfmeier, K. (2014). Forecast evaluation of an observing system simulation experiment assimilating both radar and satellite data. *Monthly Weather Review*, *142*(1), 107–124.

Kalnay, E. (2002, November 7). *Atmospheric modeling, data assimilation and predictability* [Higher education from cambridge university press].

Karspeck, A. R., & Anderson, J. L. (2007). Experimental implementation of an ensemble adjustment filter for an intermediate ENSO model. *Journal of Climate*, *20*(18), 4638–4658.

Kershaw, H. (2021). *Precomputed forward observations (FOs)* [GitHub]. Retrieved February 14, 2024, from https://github.com/NCAR/DART/wiki/Precomputed-Forward-Observations-(FOs)

Kugler, L., Anderson, J. L., & Weissmann, M. (2023). Potential impact of all-sky assimilation of visible and infrared satellite observations compared with radar reflectivity for convective-scale numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *149*(757), 3623–3644.

Kurzrock, F., Cros, S., Chane-Ming, F., Otkin, J., Hutt, A., Linguet, L., Lajoie, G., & Potthast, R. (2018). A review of the use of geostationary satellite observations in regional-scale models for short-term cloud forecasting. *Meteorologische Zeitschrift*, *27*(4), 277.

Markowski, P., & Richardson, Y. (2010, February 5). *Mesoscale meteorology in midlatitudes*. John Wiley & Sons, Ltd.

Masutani, M., Schlatter, T. W., Errico, R. M., Stoffelen, A., Andersson, E., Lahoz, W., Woollen, J. S., Emmitt, G. D., Riishøjgaard, L., & Lord, S. J. (2010). Observing system simulation experiments. In W. Lahoz, B. Khattatov & R. Menard (Eds.), *Data assimilation: Making sense of observations* (pp. 647–679). Springer.

McMurdie, W. L., & Houze, R. A. (2006, January 1). 8 - weather systems. In J. M. Wallace & P. V. Hobbs (Eds.), *Atmospheric science (second edition)* (pp. 313–373). Academic Press.

NCAR. (2009). *DART LAB - tutorial, section 1*. Retrieved February 13, 2024, from https://dart.ucar.edu/pdfs/dart_lab/DART_LAB_Section01.pdf

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., & Serafin, S. (2023, August 10). *The fractions skill score for ensemble forecast verification* (preprint). Preprints.

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.

Orlanski, I. (1975). A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, *56*(5), 527–530.

Pu, Z., & Kalnay, E. (2018). Numerical weather prediction basics: Models, numerical methods, and data assimilation. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke & J. C. Schaake (Eds.), *Handbook of hydrometeorological ensemble forecasting* (pp. 1–31). Springer.

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1), 78–97.

Roebber, P. J., Schultz, D. M., Colle, B. A., & Stensrud, D. J. (2004). Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Weather and Forecasting*, *19*(5), 936–949.

Rossa, A., Nurmi, P., & Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In S. Michaelides (Ed.), *Precipitation: Advances in measurement, estimation and prediction* (pp. 419–452). Springer.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N., & Lupu, C. (2018). An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, *11*(7), 2717–2737.

Scheck, L., Frèrebeau, P., Buras-Schnell, R., & Mayer, B. (2016). A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *175*, 54–67.

Scheck, L., Weissmann, M., & Bach, L. (2020). Assimilating visible satellite images for convective-scale numerical weather prediction: A case-study. *Quarterly Journal of the Royal Meteorological Society*, *146*(732), 3165–3186.

Schomburg, A., & Bach, L. (2023). *Operationelles NWV-system änderungsmitteilung*. Retrieved February 14, 2024, from https://www.dwd.de/DE/fachnutzer/forschung_lehre/numerische_wettervorhersage/nwv_aenderungen/_functions/DownloadBox_modellaenderungen/icon_d2/pdf_2023/pdf_icon_d2_15_01_2023.pdf?__blob=publicationFile&v=2

Schröttle, J., Weissmann, M., Scheck, L., & Hutt, A. (2020). Assimilating visible and infrared radiances in idealized simulations of deep convection. *Monthly Weather Review*, *148*(11), 4357–4375.

Skamarock. (2004). Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, *132*(12), 3019–3032.

Skamarock, Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., & Huang, X.-y. (2021). A description of the advanced research WRF model version 4.3.

Sun, J., & Wang, H. (2013). WRF-ARW variational storm-scale data assimilation: Current capabilities and future developments. *Advances in Meteorology*, *2013*, e815910.

Trapp, R. J. (2013, March 25). *Mesoscale-convective processes in the atmosphere* (1st ed.). Cambridge University Press.

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089.

Wolfgruber, L. (2024, January 16). *Wolfgruber/FSSprob: FSSprob: Computing the fractions skill score (FSS) for ensemble forecasts.* (Version v0.2.1). Zenodo.

Ying, Y. (2019). A multiscale alignment method for ensemble filtering with displacement errors. *Monthly Weather Review*, *147*(12), 4553–4565.

Ying, Y., Anderson, J. L., & Bertino, L. (2023). Improving vortex position accuracy with a new multiscale alignment ensemble filter. *Monthly Weather Review*, *151*(6), 1387–1405.

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., & Emanuel, K. (2019). What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, *76*(4), 1077–1091.