Combined State-Parameter Estimation with the LETKF for Convective-Scale Weather Forecasting

Y. RUCKSTUHL AND T. JANJIĆ

Meteorological Institute Munich, Ludwig-Maximilians-Universität München, Munich, Germany

(Manuscript received 10 July 2019, in final form 5 December 2019)

ABSTRACT

We investigate the feasibility of addressing model error by perturbing and estimating uncertain static model parameters using the localized ensemble transform Kalman filter. In particular we use the augmented state approach, where parameters are updated by observations via their correlation with observed state variables. This online approach offers a flexible, yet consistent way to better fit model variables affected by the chosen parameters to observations, while ensuring feasible model states. We show in a nearly operational convection-permitting configuration that the prediction of clouds and precipitation with the COSMO-DE model is improved if the two-dimensional roughness length parameter is estimated with the augmented state approach. Here, the targeted model error is the roughness length itself and the surface fluxes, which influence the initiation of convection. At analysis time, Gaussian noise with a specified correlation matrix is added to the roughness length to regulate the parameter spread. In the northern part of the COSMO-DE domain, where the terrain is mostly flat and assimilated surface wind measurements are dense, estimating the roughness length led to improved forecasts of up to 6 h of clouds and precipitation. In the southern part of the domain, the parameter estimation was detrimental unless the correlation length scale of the Gaussian noise that is added to the roughness length is increased. The impact of the parameter estimation was found to be larger when synoptic forcing is weak and the model output is more sensitive to the roughness length.

1. Introduction

Due to the chaotic nature of convection, ensemble forecasts are employed operationally to interpret the reliability of any deterministic convective-scale weather forecast. Aside from tracking the uncertainty of a forecast, correct representation of the error statistics of the background is crucial to determine the optimal contribution of the sparse observations to the initial conditions (analysis). To account for the highly flow dependent nature of the error statistics and to generate a statistically consistent analysis ensemble, ensemble filters have become popular. However, operationally used ensemble filters can be degraded by unrepresented model error. Model error in this work is defined in its totality and could contain both the mean and random components.

Model equations are often simplified and discretized in time and space. As a result, many relevant physical processes are not resolved and their mean impact on the model equations needs to be parameterized. These parameterization schemes typically contain several tunable parameters that are either not physical or only crudely known, leading to model errors. Since these model errors contribute to the background error, they should be represented in the background ensemble. However, as the model error is insufficiently known, it is either ignored or modeled by ad hoc techniques.

Several successful stochastic physical parameterization schemes have been developed with the aim to account for the variability in addition to the mean of the impact of unresolved processes on the model equations. Examples are stochastically perturbed parameterization tendencies (SPPT; Buizza et al. 1999), the multiphysics approach (Murphy et al. 2004), the stochastic kinetic energy backscatter (SKEB; Shutts 2005), and the physically based stochastic perturbation scheme (PSP; Kober and Craig 2016; Rasp et al. 2018; Hirt et al. 2019). However, in operational systems these schemes on their

Denotes content that is immediately available upon publication as open access.

Corresponding author: Yvonne Ruckstuhl, yvonne.ruckstuhl@lmu.de

DOI: 10.1175/MWR-D-19-0233.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

own are not sufficient to fully avoid under dispersion of the background and, by extension, ensemble forecasts (e.g., Berner et al. 2011; Zeng et al. 2019a, manuscript submitted to *Mon. Wea. Rev.*). As overconfidence of the background results in suboptimal use of observations that can lead to filter divergence (Dee 1995), it remains necessary to address model error in the data assimilation framework as well.

To complement any use of stochastic physical parameterization schemes, model error in data assimilation is typically accounted for by necessary inflation techniques embedded in the data assimilation algorithm. Perhaps the simplest form is multiplicative inflation (Anderson and Anderson 1999), where the error covariance matrix of the background $\mathbf{P}^b$ is multiplied with a factor $\rho > 1$. Such covariance inflation techniques are helpful, but they are not designed to target specific model errors, leading to suboptimal results. As an alternative, several successful additive inflation approaches have been proposed, where model error samples are added to the initial conditions in a stochastic manner. These samples are added to each ensemble member and can be obtained from climatological systems with the aim to account for large-scale errors (Rhodin et al. 2013; Zängl et al. 2015; Zeng et al. 2018), or from a higher-resolution system to account for subgrid-scale errors (Zeng et al. 2019b). Sommer and Janjić (2018) proposed augmenting the background ensemble with similar samples instead of adding them to the posterior ensemble, thereby creating a large synthetic ensemble at affordable cost to improve the quality of $\mathbf{P}^b$. They show in an idealized setup that this method reduces the need for multiplicative inflation and localization.

Additive inflation techniques, though effective, usually do not directly consider observational data, which is a rich source of flow dependent information. Recently Zeng et al. (2019a, manuscript submitted to *Mon. Wea. Rev.*) showed that discarded radar data can be used to treat a part of model error by triggering warm bubbles in appropriate locations to initiate convective cells that are observed by radars but missing in the background ensemble. They found this approach effective at reducing the precipitation error for forecasts up to three hours. Observational information can also be used in a statistical sense, by perturbing and estimating uncertain model parameters during data assimilation with the augmented state approach (Jazwinski 1970; Evensen 2009). As motivated above, parameterization schemes contain parameters that are supposed to summarize the behavior of several subgrid-scale physical processes. For example, the two-dimensional roughness length parameter is defined as the height above the surface at

which the mean wind becomes zero when extrapolating the logarithmic wind speed profile downward. As only one value for the roughness length can be assigned to each grid point, the effects of the potentially highly variable land-use and subgrid-scale orography need to be summarized for each grid box, which cannot be done without introducing model error. Many studies show that in low-resolution models, under ideal conditions (i.e., conditions in which the only model error source is the estimated parameter), the joint state and parameter estimation approach has the potential to significantly improve forecasts (e.g., Aksoy et al. 2006; Koyama and Watanabe 2010; Schirber et al. 2013; Ruiz et al. 2013). However, it is not clear how the joint state and parameter estimation algorithm reacts to the highly nonlinear relations and non-Gaussian error statistics inherited from convection. Nor is it clear how the parameter estimation is affected by model errors that are not directly related to the parameter.

Ruiz and Pulido (2015) took a step toward operational complexity in a simple atmospheric circulation model by producing an imperfect model in which some parameters are modified from original value and only a subsection of them are estimated. In presence of such simulated model error, though the estimated parameters did not converge to their corresponding perfect model value, the forecast skill did improve by compensating for the errors introduced by the perturbed parameters that were not estimated. In the presence of unknown and correlated model errors it is even harder to predict how parameter estimation impacts state predictions. A few studies have used real-world observations to explore the potential of parameter estimation for operational use (Annan et al. 2005; Kondrashov et al. 2008; Schirber et al. 2013; Doron et al. 2013; Simon et al. 2015; Kotsuki et al. 2018). All of these studies were conducted on large-scale atmospheric or ocean models. The overarching conclusion of these studies is that forecasts generally benefit from parameter estimation on short time scales, but in some cases biases are introduced on climatological scales. In addition to experiments with the assimilation of real-world observations, some of these studies conducted corresponding idealized experiments where different forms of model error were introduced to investigate the accuracy of the parameter estimation under varying conditions. They all found the same as Ruiz and Pulido (2015): parameters do not converge to their true value in the presence of model error, even if they do in a perfect model setting.

These findings are not surprising, as the augmented state approach updates parameters in such a way that the full model state is optimized with respect to the observations. The estimated parameters are therefore

used to compensate for other, possibly flow dependent model error as well. As the afore mentioned studies indicate, this can be a disadvantage when one is interested in retrieving a converged temporally constant parameter value to use for a long-term forecast, since the affected model error acts on time scales smaller than the forecast lead time. On the other hand, the same studies suggest that forecasts of state variables can be improved by allowing the estimated parameter to compensate for model error that acts on larger time scales than the forecast lead time. This was, for example, emphasized by [Koyama and Watanabe \(2010\)](#), where estimation of temporally varying parameters was found to be beneficial for limited forecast lead times only. For longer forecasts, starting from the climatological parameter value rather than (close to) the exact parameter value at analysis time was found to yield better results. Since in this paper we are interested in convective-scale numerical weather prediction for short-term forecasts up to 6 h, we do not attempt to restrict the scope affected model error. Instead, we rely on available observations, the quality of the numerical prediction model and the data assimilation algorithm to provide a parameter evolution that leads to improved forecasts of the state variables. Note that this perspective implies that the estimated parameter is considered flow dependent regardless of its physical meaning. We justify this further by claiming that the effect of model error stemming from a temporally constant physical parameter could be flow dependent. For example, when a spatially varying parameter is discretized in space, like the roughness length, the unrepresented subgrid-scale variability can have a flow dependent effect on state variables. In this case we cannot assume that a temporally constant “correct” parameter projected on the numerical grid even exists.

In this paper we explore the feasibility of improving convection-permitting numerical weather prediction by estimating uncertain model parameters using an ensemble filter in nearly operational configuration. Our experiments are conducted with the Consortium for Small-Scale Modeling (COSMO; [Baldauf et al. 2011](#)) model and corresponding data assimilation system Kilometer-Scale Ensemble-based Data Assimilation (KENDA; [Schraff et al. 2016](#)), which is based on the localized ensemble transform Kalman filter (LETKF; [Hunt et al. 2007](#)). The model parameter of interest is the two-dimensional roughness length z_0 , which influences convection via parameterized surface fluxes. The most direct available observations of z_0 are wind measurements close to Earth’s surface. However, since z_0 influences wind only via surface fluxes, the algorithm cannot distinguish model errors caused by the roughness length

from those caused by surface fluxes. We therefore aim for a flow dependent parameter that addresses both sources of model error. We thereby hope to improve the simulation of convective clouds in the COSMO model, which we verify by comparing the resulting model forecasts to independent visible and near-infrared images of Meteorological Satellite (Meteosat) Spinning Enhanced Visible and Infrared Imager (SEVIRI) observations.

In [section 2](#) we introduce the COSMO model and explain the role of the roughness length in the context of the model equations. Then we discuss the observations that are assimilated and used for verification in [section 3](#). In [section 4](#) we explain our experiment setup, which includes a description of the data assimilation system, specifications regarding the parameter estimation and a brief overview of the different weather regimes that marked our test period. The model equivalents (i.e., model estimates of the true state of the atmosphere in observation space) of our experiments are verified against Meteosat SEVIRI observations and radar reflectivity to measure the quality of the representation of clouds and precipitation respectively. The results are analyzed in [section 5](#), followed by a conclusion in [section 6](#). This paper is completed with a critical discussion of possible improvements and suggestions for future work in [section 7](#).

2. COSMO

The COSMO-Model is a nonhydrostatic limited-area numerical atmospheric prediction model that is developed and maintained by the national weather services of the COSMO consortium for operational weather forecasting as well as research. Its dynamical core is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere, derived from the basic conservation laws for mass, heat, momentum and energy. Its practical formulation consists of a set of prognostic equations describing wind, pressure perturbation, temperature, total air density and mass fraction of water in different phases. We employ the convective-scale configuration COSMO-DE ([Baldauf et al. 2011](#)), which covers Germany and parts of neighboring countries. The horizontal grid spacing is 2.8 km, allowing deep convection to be resolved partially. The 50 hybrid vertical layers extend from the surface to a height of 22 km (~ 40 hPa) and gradually shift from terrain following at the bottom to horizontally flat at the top. Lateral boundary conditions are provided twice a day by the prediction system of the global Icosahedral Nonhydrostatic (ICON; [Zängl et al. 2015](#)) model, which runs globally with a resolution of 40 km

for ensemble members and 20 km for the deterministic run. Over Europe the resolution is increased to 13 and 6.5 km, respectively. The boundary conditions at the ground are provided by the multilayer soil model TERRA (Doms et al. 2011).

While the dynamical core equations are capable of representing key aspects of atmospheric flow, including deep convection in part, the numerical grid is too coarse to resolve all relevant processes explicitly. In particular, shallow convection is parameterized based on the Tiedtke scheme (Tiedtke 1989), clouds are represented using a convection-allowing Lin–Farley–Orville-type one moment bulk microphysical scheme, including cloud droplets, cloud ice, rain, snow and graupel (Lin et al. 1983; Reinhardt and Seifert 2006), the heating rate caused by radiation is computed according to Ritter and Geleyn (1992), subgrid-scale turbulence is parameterized based on the turbulent kinetic energy (TKE) equation described in Raschendorfer (2001) using parameterized surface fluxes as lower bound. The roughness length influences the dynamical flow via these surface fluxes.

The representation of the roughness length z_0 in COSMO depends on subgrid-scale orography $z_{0,\text{or}}$ and land use $z_{0,\text{lu}}$. Only the mean and variance of subgrid-scale orography is taken into account, $z_{0,\text{or}} = a_0 \sigma^2 \arctan(\Delta x / b_0)$, where σ^2 is the variance of the subgrid-scale orography scaled by its mean, $a_0 = 10^{-5}$ m, $b_0 = 2.5$ m, and $\Delta x = 2.8$ km. The roughness lengths $z_{0,i}$ corresponding to the different land uses i (grass, crops, forest, urban, etc.) within one grid box of area A , are logarithmically averaged weighted by their respective areas A_i :

$$z_{0,\text{lu}} = h \exp \left[- \frac{A}{\sum_{i=1}^I A_i \ln \left(\frac{h}{z_{0,i}} \right)} \right], \quad (1)$$

where h is the height of the lowest grid level above the ground. The final roughness length value used in COSMO is the sum of the two contributions:

$$z_0 = z_{0,\text{or}} + z_{0,\text{lu}}. \quad (2)$$

To avoid excessive heat exchange over rough terrain, the roughness length for heat $z_h = \max(z_0, z_{h,\text{max}})$ is introduced, which is used to calculate the surface heat and vapor fluxes. We use the operational value of $z_{h,\text{max}} = 0.98$.

3. Observations

Data assimilation system KENDA assimilates conventional observations and radar reflectivity for regional

weather predictions (Schraff et al. 2016; Gustafsson et al. 2018). In addition, in this study visible and near-infrared satellite images from Meteosat SEVIRI are used for model output verification purposes. All observations undergo quality control as done operationally by the German Weather Service (Schraff and Hess 2012).

a. Conventional observations

Conventional observations measure prognostic variables of COSMO, such as temperature, wind, pressure and relative humidity. Synoptic surface data (SYNOP) provides hourly measurements of surface pressure, horizontal wind at 10 m above the surface, and temperature and humidity at 2 m above the surface. Surface wind measurements are only consistently assimilated in the north of Germany. In the south they are usually rejected because the model cannot adequately resolve the surface height due to subgrid-scale orography, causing the observations to differ too much from the interpolated model output. Upper air data such as radiosondes (TEMP), wind profilers (PROF), and aircraft reports (AIREP) do not share this problem and are therefore assimilated when available. Radiosondes are typically launched at 0000 and 1200 UTC with some in addition launching at 0600 and 1800 UTC. PROF measurements of horizontal wind are available approximately half-hourly during the whole day and the majority of AIREP horizontal wind and temperature data are collected during the daytime. These include wind and temperature observations obtained from aircraft within the range of airport radars in the surveillance Mode-S EHS (Lange and Janjić 2016).

b. Radar data

Aside from conventional observations, 17 C-band Doppler radars over the COSMO-DE domain provide reflectivity and radial wind measurements every 5 min. In this study, like conventional observations, we assimilate radar reflectivity directly with the LETKF using an efficient radar forward operator EMVORADO (Zeng et al. 2014, 2016) that converts a model state to its reflectivity equivalent. Experiments conducted by Bick et al. (2016) and Zeng et al. (2018, 2019b) indicate high potential for assimilation of radar reflectivity with the LETKF. They found that temporal thinning is necessary to reduce temporally correlated observation errors. Therefore only the latest 5 min of available reflectivity measurements are assimilated. In addition, the same spatial superobbing is applied as in Bick et al. (2016) and Zeng et al. (2018, 2019b). As in Aksoy et al. (2009), Bick et al. (2016), and Zeng et al. (2018, 2019b) reflectivities smaller than 5 dBZ are assimilated as no-reflectivity information.

c. Satellite data

Visible satellite images obtained from the Meteosat SEVIRI instrument contain high-resolution information about clouds, which can be used for data assimilation purposes as well as model output verification. For this application, [Scheck et al. \(2016, 2018\)](#) recently proposed the forward operator based on the Method for Fast Satellite Image Synthesis (MFASIS) to generate synthetic images from the COSMO-DE output that is fast and compact enough for operational use. Visible satellite images are available every 5 min over Europe and every 15 min elsewhere. Since applying this forward operator to assimilate visible satellite images is currently in research mode, we here use it for verification purposes only.

4. Experiment setup

The data assimilation system KENDA is based on the LETKF ([Hunt et al. 2007](#)) for assimilation of conventional data. To deal with undersampling, model and observation errors and computational limits, different techniques are applied in addition to the standard LETKF equations. In this section we specify the KENDA data assimilation settings that we used and radar data assimilation settings for the LETKF. In addition, we introduce the dynamical model for the roughness length and provide some background on the weather situation of the test period used for this study.

a. Data assimilation

The experiments are conducted in a nearly operational setting using a basic cycling environment (BACY) developed at the German Weather Service. We focus on the period from 28 May to 9 June in 2016, which we split into two periods of 6 days, due to the weather situation (see [section 4c](#)). We use a 40 member ensemble for the hourly cycling. Every hour from 0600 to 1500 UTC we start a 6 h forecast using 20 ensemble members, so that for each of the two weeks we have a total of 60 forecasts of 20 members available for verification. Note that our choice of forecast start times stems from the availability of visible satellite data.

The LETKF updates the prognostic variables of TKE, the mass fraction of water in its different phases, the three-dimensional wind components, temperature, pressure perturbation, specific humidity, cloud water and ice hourly. In KENDA, domain localization (e.g., [Hunt et al. 2007](#); [Janjić et al. 2011](#)) is done with the weighting of observations using the Gaspari–Cohn function ([Gaspari and Cohn 1999](#)). The localization radius is set per grid point such that 100 observations

are assimilated, subject to a lower and upper limit of 50 and 100 km, respectively. For a graphical view of the resulting horizontal localization radii we refer to [Lange and Janjić \(2016\)](#). Due to the high and constant density of radar data, a fixed localization radius of 16 km is set for this particular observation type. The vertical localization ranges from 0.075 (surface) to 0.5 (top) in logarithm of pressure. As a result, the roughness length can be influenced by observations up to an altitude of order 1 km.

Motivated by [Zeng et al. \(2018\)](#), the only inflation technique to deal with underdispersion of the background error covariance matrix used in this study is additive inflation. Here model error samples obtained from climatological systems are used to represent large-scale model errors. These samples are added a posteriori to the analysis ensemble ([Rhodin et al. 2013](#)). The intended effect is the increase of spread in a directed manner.

Observation errors are assumed temporally and spatially uncorrelated. The diagonal values of the observation error covariance matrix $\mathbf{R}$ based on Desroziers statistics ([Desroziers et al. 2005a,b](#)) for surface pressure and those assumed for upper air measurements are given in [Schraff et al. \(2016\)](#). For radar reflectivity a fixed error of 10 dBZ is assumed.

The KENDA system calculates the analysis weights on a coarser grid and interpolates coarse analysis weights onto the high-resolution grid to reduce computational costs ([Yang et al. 2009](#)). Without significant loss of accuracy, a horizontal coarsening factor of 3 is applied and the number of vertical layers is reduced from 50 to 30, leaving only 6.67% of the grid points to be analyzed. Furthermore, techniques like saturation adjustment and hydrostatic balancing are applied to the computed analysis, to ensure well balanced initial conditions ([Rhodin et al. 2013](#)).

b. Dynamical model for roughness length

For estimation of the roughness length z_0 we use the augmented state approach ([Jazwinski 1970](#); [Evensen 2009](#)). Since z_0 is spatially varying, horizontal and vertical localization is applied in exactly the same way as for the surface state variables. Due to the static design of z_0 in COSMO, a model propagation cannot provide any direct information about the uncertainty of z_0 . Instead, we have to explicitly model the uncertainty as a heuristic exercise. We choose the following approach, based on for example [Gelb \(1974\)](#) and [Ruckstuhl and Janjić \(2018\)](#). For each ensemble member i , $i = 1, \dots, N_{\text{ens}}$ at time t :

$$\tilde{z}_{0,i}^a = z_{0,i}^a + \mathbf{D}_t \mathbf{C}^{1/2} \boldsymbol{\eta}_t, \quad (3)$$

$$z_{0,t+1,i}^f = \tilde{z}_{0,t,i}^a, \quad (4)$$

where $z_{0,t,i}^a$ is the raw analysis value of the roughness length after applying the LETKF, $\tilde{z}_{0,t,i}^f$ the perturbed value that is passed to the model and $z_{0,t+1,i}^f$ the background value corresponding to the next cycle, $\mathbf{D}_t$ is a diagonal matrix that locally controls the ensemble spread of the roughness length, $\mathbf{C}$ is the error correlation matrix that specifies the correlations within the roughness length field, and $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{I})$ determines the random realization of the stochastic model. In contrast to relaxation or multiplicative inflation methods, the model as described by (3) adds fresh spatial structures to the roughness length perturbations, which minimizes the dependency on the initial (random) realization $\boldsymbol{\eta}_0$.

Ideally this stochastic uncertainty model should take into account all targeted unrepresented model errors, in this case the roughness length itself and the surface fluxes. As there is no evident way to estimate this accumulated uncertainty, we follow Aksoy et al. (2006) and Ruckstuhl and Janjić (2018) and choose $\mathbf{D}_t$ such that for each grid point j the spread of $\{\tilde{z}_{0,t,i}^a(j); i = 1, \dots, N_{\text{ens}}\}$ equals 25% of the corresponding original roughness length value $z_{0,t}(j)$ as implemented in COSMO-DE. Note that this choice implies that the spread of z_0 is fixed in time. In particular, if the roughness length is not affected by any observations at grid point j during the data assimilation, the spread of $\{\tilde{z}_{0,t,i}^a(j), i = 1, \dots, N_{\text{ens}}\}$ remains equal to 25% of the initial $z_{0,0}(j)$ value and so $\mathbf{D}_t(j, j) = 0$. It should be stressed that $\mathbf{D}_t$ is a tunable matrix and may affect the experiments significantly, as it determines the variance of the background error of the roughness length, which directly affects the size of the analysis increments. This is further discussed in section 7. We set $\mathbf{C}$ such that the correlations are an approximate Gaussian function of the spatial distance between roughness length elements. To test the effect of the assumed correlation length scale dx of the model error, we respectively do experiments with $dx = 0$, $dx = 5$, and $dx = 25$ grid points. Note that this model is only applied at each analysis step, so that for each ensemble member the roughness length remains constant throughout a model forecast as indicated by (4).

The choice of our configuration then implies that we assume that the accumulated targeted model error projected onto the roughness length is Gaussian with a correlation length scale of either 0, 5, or 25 grid points (where 5 grid points is the effective resolution of the model and simulates the effect of unresolved processes only), and at each grid point a temporally constant standard deviation of 25% of the climatological mean, which we assume is the original roughness length parameter

value. As the roughness length cannot be negative we set $z_{0,t,i}^a \leftarrow \max(z_{0,t,i}^a, z_0^{\min_b})$ before the dynamical model is applied and set $\tilde{z}_{0,t,i}^a \leftarrow \max(\tilde{z}_{0,t,i}^a, z_0^{\min_a})$ after the dynamical model is applied, where $0 < z_0^{\min_a} < z_0^{\min_b}$ to guarantee a spread larger than zero. We choose $z_0^{\min_a} = 0.0001$ and $z_0^{\min_b} = 0.0002$.

We run a set of 5 experiments: a reference run referred to as *ref*, where z_0 is not estimated, $dx = 0$, $dx = 5$, and $dx = 25$ referring to experiments where z_0 is estimated using the dynamical model for the respective correlation length scales, and, to determine whether any change with respect to *ref* is the result of parameter estimation or simply the addition of noise, a final experiment is conducted referred to as *perturb*, where z_0 is perturbed according to the dynamical model with correlation length scale $dx = 5$, but the mean value is left unchanged (i.e., z_0 is not estimated).

c. Weather situation

Our experiments are conducted for subsets of the two-week period from 27 May to 9 June 2016, when a large part of Europe was affected by a sequence of severe convective storms. These storms were the result of the combination of high moisture content, low thermal stability, weak wind speed, and large-scale lifting caused by surface pressure lows. These conditions persisted for two weeks due to an atmospheric blocking in the form of a large-scale ridge spreading over the North Atlantic and northern Europe, hampering mass exchange. Western and southern Germany remained under the influence of low pressure with moist and warm air, while drier air gradually prevailed in the northeast. During the first week of this blocking event, moist and warm air was advected ahead of a deep trough northeastwards toward central Europe, causing synoptic lifting. During the second week moisture was maintained mainly by evapotranspiration from local sources and advection from nearby countries. Due to the weak pressure gradient, horizontal flow was limited, causing the convective storms to be almost stationary (Piper et al. 2016), especially in the second week. The majority of the convective events during the whole period followed a typical diurnal cycle with peaks in the late afternoon (Zeng et al. 2018).

We distinguish between synoptic regimes that occurred throughout this convectively active period of two weeks by applying the convective adjustment time-scale measure (τ_c) introduced by Done et al. (2006) and Keil et al. (2014). Analogous to Keil et al. (2014), we define a day to be weakly synoptically forced if τ_c exceeds the threshold of 6 h at least once, and strongly synoptically forced otherwise. This leads to a natural split of our experiments in two 7 days periods. From 27 May 2016

FIG. 1. Snapshot of gridpoint-wise correlation between z_0 and the u component of the wind 10 m above the surface for (left) $dx = 0$, (middle) $dx = 5$, and (right) $dx = 25$.

to 2 June 2016 the synoptic forcing is strong, and from 3 June 2016 to 9 June 2016 the forcing is weak, with the exception of 9 June 2016 (see Zeng et al. 2018). For each week, the first day is used for spinup, leaving 6 days for evaluation.

5. Results

As stronger synoptic forcing suggests weaker influence from subgrid-scale processes, we hypothesize that our approach is beneficial for the second week, and neutral for the first week of our test period. We therefore focus on the experiments conducted in the weakly forced weather regime, from 4 June to 9 June 2016 first.

As horizontal wind at 10 m above the surface is strongly correlated to the roughness length (see Fig. 1), we expect wind measurements near the surface to be the most important observations for the estimation of z_0 . Figure 2 visualizes the spatial distribution of all wind observations assimilated. It is clear that south from 50°N the representation of near-surface wind measurements is very poor. For evaluation of the results we therefore divide the domain into north (above 50°N latitude), where surface wind measurements are mostly assimilated, and south (below 50°N latitude), where surface wind measurements are mostly discarded due to quality control. We further justify this split by pointing out that the higher variability of the orography in the south (for

FIG. 2. (left) Domain extent of COSMO-DE with conventional wind observations of types PILOT (yellow), SYNOP (blue), TEMP (red), and AIREP (green) that are assimilated. The area of the circles is proportional to the average daily number of single wind observations at the indicated horizontal locations for the period from 4 to 9 Jun 2016. The average number of observations assimilated per day are indicated in the legend. (right) Scatterplot featuring the latitude and height (surface to the pressure level of 850 hPa) of all conventional wind observations assimilated on a representative day (4 Jun). Note that 850 hPa is approximately the height where data directly influence the roughness length due to vertical localization.

FIG. 3. Snapshot (1000 UTC 9 Jun) of the deterministic roughness length for *ref*, $dx = 0$, $dx = 5$, and $dx = 25$.

instance due to the Alps), is an important source of model error, which the roughness length could potentially (partially) compensate for. However, without surface wind measurements the roughness length is left to be influenced by observations of quantities that are correlated to the roughness length via a more complex causal chain. As a result, the lack of surface wind observations in combination with the large model error inherited from orography, causes the behavior of the parameter estimation to be especially unpredictable in the south. See for example [Schraff et al. \(2016\)](#) for visualization of the orography as represented in COSMO-DE.

We start by analyzing the changes in the roughness length parameter and then continue to verification of atmospheric fields.

a. Roughness length evolution

Figure 3 shows a snapshot of z_0 after 6 days of hourly cycling. The spatial correlation inherited from the different

dynamical models for z_0 is clearly visible. For all experiments the orography, especially the Alps, remains featured and is even emphasized. In [Fig. 4](#) the spatial mean of the absolute analysis increments for the different experiments is shown. The evolution of the mean increments clearly highlights a diurnal cycle. A natural assumption is that this is related to the diurnal cycle of the surface fluxes. Indeed, the spatial mean of the hourly differences of the momentum surface flux (dashed black line) evolves in sync with the z_0 increments. The surface heat and vapor fluxes generally peak earlier in the day (not shown). It is encouraging that the increments appear to be driven by physical processes rather than spurious correlations, at least during the day. That said, spurious correlations do influence the evolution of z_0 , as can be interpreted from the noisy nature of the increments (dotted lines in [Fig. 4](#)). Also, the increments in the south are roughly twice as large as in the north, most likely due to the choice of the dynamical model for the

FIG. 4. Absolute analysis increments averaged over space for (top) the north and (bottom) the south. The increments are separated into a smoothed part (solid lines) and the remaining noise (dotted lines) using the Savitzky-Golay filter. The dashed black line is the average hourly absolute surface momentum flux differences corresponding to the reference run, normalized with some linear function for visualization purposes.

roughness length. Indeed, recall that the parameter spread is proportional to the parameter value. Also, numerical weather prediction systems have difficulties providing accurate forecasts over mountainous terrain, because the effect of subgrid-scale orography is large and difficult to model. The model error is therefore expected to be large over the Alps, which the LETKF tries to compensate for by modifying the roughness length, leading to large increments.

From Fig. 4 we cannot establish a clear difference in behavior for the three dynamical models. Even histograms of the absolute increments (not shown), do not provide information on any distinction. However, when looking at the spatial distribution of the analysis

increments, clear differences are found, see Fig. 5. As expected, the smaller the correlation length scale of the dynamical model, the “spottier” the increments. This effect, though heavily reduced, is passed on to the surface momentum flux (Fig. 6), where $dx = 25$ mainly accentuates the existing features of the reference run, $dx = 5$ introduces lighter and darker spots of similar size as seen in Fig. 5, and $dx = 0$ clearly introduces higher-resolution features. Yet, the differences appearing in the surface momentum flux among the different dynamical models are subtle with respect to those in the roughness length (Fig. 3). Despite the clear differences in the roughness length, the main features of the surface fluxes remain similar to the reference run for all dynamical

FIG. 5. Snapshot (1000 UTC 9 Jun) of analysis increments for (left) $dx = 0$, (middle) $dx = 5$, and (right) $dx = 25$.

FIG. 6. Zoomed-in snapshot (1000 UTC 9 Jun) of surface momentum flux (N m^{-2}) generated from the deterministic background for the different experiments: (clockwise from top left) *ref*, $dx = 0$, $dx = 5$, and $dx = 25$.

models. The differences for the heat and vapor fluxes are even smaller (not shown), most likely due to the upper bound imposed on the roughness length for heat.

The statistical properties of the spatial distribution of the roughness length shown in Fig. 7 highlights steady behavior in the north and more noisy behavior in the south, especially for $dx = 25$. In general $dx = 0$ and $dx = 5$ exhibit similar behavior. In both regions the standard deviation is steadily increasing, while the median appears to drop somewhat. The mean increases slightly, which is partially due to the imposed lower bound of 0.0002 m. These features seem more pronounced in the south. For $dx = 25$ however, the south exhibits different behavior than in the north. In the north all statistical properties shown are higher than for the other experiments, which is not true in the south. An important note is that $dx = 25$ introduces a large bias with respect to the original z_0 in the north. Whether this is the result of an existing model error bias that should be corrected, or simply an imposed effect of the chosen dynamical model for z_0 , can be clarified using the forecast scores discussed in a later section. A matter of concern is that the evolution of the roughness length does not saturate during our test period. Recall that we

do not need the parameter to converge, but we do want to keep the parameter within reasonable bounds. We revisit this issue in section 7.

In general it seems the three dynamical models show very different behavior in terms of the spatial distribution of z_0 , yet show very similar behavior in terms of average change per cycle. Figure 7 suggests the sign of the increments differs among the experiments: a large correlation length scale appears to result in larger positive than negative increments, introducing a bias with respect to the original parameter.

b. Verification of atmospheric fields against observations

We first look at the average RMSE and spread of the experiments with respect to assimilated observations for the background, see Fig. 8. What is most notable, especially in the north (south is not shown), is that *perturb* has by far the largest overall RMSE of all experiments, including the control. From the right plot in Fig. 8 we can also conclude that randomly adding perturbations does not necessarily result in a larger spread. However, when the perturbations of z_0 are used to allow change in the mean of z_0 , the spread increases and the RMSE

FIG. 7. Evolution of statistical properties of the spatial distribution of z_0 corresponding to the deterministic run for the cycling period. The mean (solid), standard deviation (dashed), and median (dotted) are plotted for (top) the north and (bottom) the south. Colors indicate the different experiments.

decreases. This is not surprising since the LETKF seeks to minimize the distance between the first guess and the assimilated observations, using the additional degrees of freedom offered by the variable parameter for that purpose. However, initial conditions that are closer to the observations, do not guarantee they remain closer throughout a forecast and vice versa. We are interested in longer forecast lead times than 1 h. Also, the

quantities we are most interested in are clouds and precipitation, as they describe convective events. These we verify next.

1) REFLECTANCE

Since Meteosat SEVIRI observations are not reliable over the Alps due to the possibility of snow, the domain is cut off at a latitude of approximately 47.5° . [Figure 9](#)

FIG. 8. (left) Background RMSE compared to the assimilated conventional observations and (right) spread of $dx = 0$, $dx = 5$, and $dx = 25$ and *perturb* relative to *ref*, averaged over all assimilation cycles. Numbers indicate the total number of assimilated observations.

FIG. 9. Snapshot of SEVIRI observations (top left) and model equivalents of (left) *ref* and (right) $dx = 5$ for a 4-h forecast started at 0700 UTC 7 Jun 2016. (second row) A snapshot of the deterministic run and (bottom row) a probability plot generated from the corresponding ensemble with a threshold of 0.3.

offers a snapshot of the reflectance, which illustrates the influence that estimating z_0 can have on the representation of clouds in COSMO-DE. The differences are mostly subtle, but there are areas where a clear difference can be spotted, see for example the circled area. After estimating the roughness length, both the deterministic run and the corresponding ensemble predict clouds with a reflectance larger than 0.3. If the roughness length is not estimated as in the control experiment, neither deterministic run nor corresponding ensemble show these reflectance values in that area. More quantitative results are presented in Fig. 10 in the form of bias, RMSE, spread, and fractions skill score (FSS, Roberts and Lean 2008), averaged over all 60 forecasts. Markers in the figures indicate statistical significance according to the bootstrap method (Efron and Tibshirani 1994), where the absolute differences compared to *ref* are calculated, followed by 10 000 bootstrap resamplings to determine the statistical significance for a 95% confidence interval. As before, simply perturbing z_0 without

estimation is yielding poorer results than $dx = 5$, especially in the south. We can therefore conclude that any gain is due to updating z_0 based on correlations with observed variables, rather than being the result of adding noise. Differences between north and south are most pronounced for $dx = 0$. In the north this experiment significantly outperforms all other experiments, whereas in the south it has the worst scores. In contrast, $dx = 5$ is steadily performing better than both $dx = 25$ and the control run over the entire domain, except for the bias, where $dx = 25$ is superior. The FSS is widely used to measure the skill for different scales. In Figs. 11 and 12 we see that the FSS relative to the control run is more sensitive to the threshold than the scale. Again we see that $dx = 0$ excels in the north, especially for high thresholds and longer forecast lead times, achieving an improvement of more than 5%. We also computed the false alarm ratio (FAR; Wilks 2006) and equitable threat score (ETS; Wilks 2006) (not shown) and the results are consistent with the FSS. The FAR and the ETS

FIG. 10. (from top to bottom) Reflectance bias, RMSE, spread, and FSS of the respective experiments for the (left) entire domain, (middle) the north and (right) the south for the weakly synoptically forced week. The FSS ranges from a perfect score of 1 to lowest score of 0 and is calculated for a scale of 36 km and thresholds 0.3 and 0.5. Markers indicate statistical significance.

complement each other in the sense that the FAR is meant to provide a measure for the amount of spurious convection, whereas the ETS measures the amount correctly predicted convection. Being superior in both

scores indicates that better scores are not just the result of over or under forecasting.

To verify that surface wind measurements are important for the estimation of the roughness length, we

FIG. 11. Relative reflectance FSS score in percentage (color) with respect to *ref* in the north for the weakly synoptically forced week for the experiments (rows) for different lead times (columns).

did additional experiments where we did not assimilate any SYNOP observations. As can be concluded from Fig. 13, SYNOP data plays an important role in the parameter estimation for $dx = 0$. Without this observation type the parameter estimation is detrimental for the forecast, which is evidence that supports the need for surface wind measurements.

2) PRECIPITATION

The verification product for precipitation is a combination of radar derived precipitation measurements and rain gauges. As this product is only available to us over Germany, the verification region is masked to its borders. Due to the existence of outliers in radar data, the bias and RMSE are deemed inappropriate scores for reflectivity. Instead, scores like the FSS, FAR, and ETS are used, where the concept of thresholds eliminates the sensitivity to outliers.

From Fig. 14 we conclude that differences for the entire domain among the experiments are smaller for reflectivity scores than for reflectance scores. There are however significant differences in the north. Here it seems that $dx = 5$ is predicting precipitation more accurately than the control run. Again the ETS and FAR are consistent with the FSS (not shown). Figure 15 also shows that the gain especially holds on smaller scales and higher thresholds.

The advantage peaks at 2 to 3 h forecast lead time where the gain is over 20%. Such a significant feedback between surface quantities and precipitation was also established by for example Ha and Snyder (2014), who showed that assimilation of surface measurements improves the dynamics of the planetary boundary layer, resulting in improvements of precipitation forecasts.

As a side note, it is clear that for both the reflectivity and precipitation the scores are better in the south than in the north for all experiments. Bachmann et al. (2019) found similar results, and argue that the orography in the south enhances predictability due to its ability to trigger convection.

3) STRONG SYNOPTICALLY FORCED WEEK

Figures 16 and 17 show the verification scores for reflectance and precipitation for the strongly synoptically forced week, respectively. As we already concluded that updating the mean of z_0 is crucial to outperform the control run, we did not compute the *perturb* experiment for this period. Neither do we show the $dx = 25$ experiment, as we found that a correlation length scale of 25 grid points is too large for the dynamical model of the roughness length.

The differences among the experiments are smaller than for the weakly synoptically forced week (below 1%

FIG. 12. As in Fig. 11, but for the south.

for the FSS). This supports our hypothesis that the effect of estimating the roughness length is smaller when the synoptic forcing is larger. In the north the scores indicate a slightly improved forecast with respect to the control run, except for the bias. In contrast to the weakly synoptically forced week, $dx = 0$ is doing better than the control run in the south. An exception is again the bias, though it should be noted that the bias is very small in the first few forecast hours in comparison to the north and the weakly synoptically forced week. As the roughness length is less bounded in the south than in north due to lack of assimilated surface wind observations, it is less predictable how the LETKF utilizes the

degrees of freedom inherited from the parameter estimation.

6. Conclusions

We conclude that estimating the roughness length with the augmented state approach can lead to better predictions of clouds and precipitation. The clear diurnal cycle of the analysis increments of the roughness length indicates the influence of physically based correlations. The diurnal cycle of the surface momentum flux coincides with changes in the roughness length, while the surface heat and vapor fluxes peak earlier in

FIG. 13. (from left to right) Bias, RMSE, spread, and FSS at a scale of 36 km and thresholds 0.3 and 0.5 for the north. The experiments *ref* and $dx = 0$ (solid) are as in Fig. 10. Experiments *ref*\SYNOP and $dx = 0$ \SYNOP (dashed) are as *ref* and $dx = 0$ respectively, but without assimilation of SYNOP data.

FIG. 14. Reflectivity (top) FSS for thresholds 1 and 5 mm h⁻¹ and (bottom) the spread of the respective experiments for (left) the entire domain, (middle) the north, and (right) the south for the weakly synoptically forced week. Markers indicate the same as in Fig. 10.

the day. This suggests that the parameter estimation mostly influences the state variables via the momentum fluxes. In addition, the best forecast scores with respect to the control run (*ref*) were achieved in the north, where surface wind measurements are widely available, for a dynamical model with no correlation length scale ($dx = 0$). The importance of these wind measurements was confirmed by repeating the experiments *ref* and $dx = 0$ with exclusion of SYNOP data in the assimilation process, which led to a significant detrimental impact of parameter estimation on the forecast. This highlights the need for a healthy coverage of observations that are closely related to the estimated parameter.

In the north the scores were consistently better than the control run for a dynamical model with a modest correlation length scale ($dx = 0$ and $dx = 5$). For a larger correlation length scale ($dx = 25$) and ceasing to update the parameter mean (*perturb*) no improvement over the control run was found. This confirms that the influence of observations on the evolution of the roughness length was beneficial for the prediction of clouds and precipitation, and that imposing artificial correlations in the roughness length field should only be done modestly.

In the northern part of the domain, experiment $dx = 0$ was superior for the prediction of clouds, and $dx = 5$ was superior for the prediction of precipitation. Yet, both experiments were consistently outperforming the control run. This was especially profound in the weakly

synoptically forced week, reflecting the higher sensitivity of convection to the roughness length.

In the southern part of the domain, estimation of the roughness length with a correlation free dynamical model was clearly detrimental for the weakly synoptically forced week. For $dx = 5$ the spatial correlations imposed on the roughness length increase the influence radius of the observations assimilated. For the weakly synoptically forced week, when convection is sensitive to the roughness length, this proved favorable in the south, where model error is large and surface wind observations sparse. The choice of correlation length scale for the dynamical model of the roughness length should therefore depend on the model error and observations assimilated.

Overall, the superior results in the north support our hypothesis of the importance of assimilating surface wind measurements for the estimation of the roughness length. In the south, introducing spatial correlations within the roughness length field partially compensates for the lack of assimilated surface wind observations. Improvements are most pronounced in a weakly synoptically forced weather regime when the sensitivity of convection to the roughness length is largest.

7. Critical look and future work

The conclusions presented in section 6 are promising, but they have the potential to be much better. In this

FIG. 15. Relative reflectivity FSS score in percentage (color) with respect to *ref* for the weakly synoptically forces week in the north for the remaining experiments (rows) for different forecast lead times (columns).

section we propose some ideas on how to improve the parameter estimation.

a. Heat versus momentum flux

The roughness length for heat exchange (z_h), though heavily correlated via the landscape type, is not the same as for momentum (z_0). In fact, since momentum transfer is mainly caused by turbulent drag on roughness elements and heat exchange occurs via molecular thermal diffusion, it is necessary to add an additional resistance term to account for differences in transport mechanisms between momentum and heat (Owen and Thomson 1963; Yang and Friedl 2003). In the COSMO model the roughness length for heat z_h is set as $z_h = \min(z_0, z_{h_{\max}})$, which does not allow decoupled estimation of z_0 and z_h . Consequently, the analysis increments of z_0 are the result of a compromise between reducing the errors in the momentum flux and the surface heat flux. In addition, the imposed upper bound of z_h causes complications for data assimilation as for example discussed in Ruckstuhl and Janjić (2018). Furthermore, it has been shown that landscape variability plays an important role in the initiation of cumulus clouds via surface heat fluxes, see for example Rabin et al. (1990). However, from our results we deduce that the momentum flux dominated the parameter increments. This is possibly an effect of the upper bound z_h , which

can cause severe spread reduction in z_h . It is therefore natural to estimate the parameters z_0 and z_h separately, or at least estimate a coupling term between the two.

b. Spurious correlations

Spurious correlations are a serious problem for parameter estimation. Localization as done for the state is not necessarily appropriate for parameters, even for a spatially varying one, because the influence radius is smaller. This is especially the case for non-smooth parameters such as the roughness length. We partially addressed this issue by introducing a correlation length scale in the model, which essentially reduced the degrees of freedom. This strategy was indeed effective in the south where surface wind observations are extremely sparse. However, naturally, introducing artificial correlations is not ideal, as it may also introduce wrong correlations. This was probably the case for $dx = 25$. We believe a more natural way to reduce spurious correlations for parameter estimation is sampling error correction (SEC, Anderson 2012, 2016; Necker et al. 2020). This method is based on Monte Carlo simulations that serve to calculate a statistically based correction term for a given sample correlation, ensemble size, and prior assumption of the correlation distribution. These corrections are computed

FIG. 16. As in Fig. 10, but for the strongly synoptically forced week.

offline, resulting in a simple look up table for the sampling error correction.

Another (complementary) approach for dealing with these sampling errors is selecting a priori which type of observations is allowed to influence the roughness length. This way, observations that are believed to be

uncorrelated (or have a highly indirect relation) to the roughness length can be excluded, thereby reducing the number of spurious correlations. Kang et al. (2011) demonstrated on the problem of assimilating carbon dioxide concentrations into a dynamical model that this type of localization can be very effective.

FIG. 17. As in Fig. 14, but for the strongly synoptically forced week.

c. Tuning

Although we have explored three different correlation length scales, the dynamical model would most certainly benefit from some tuning. We saw for example large differences among experiments between the forecast quality in the north and in the south. This suggests the need for a location dependent correlation length scale [determined by $\mathbf{C}$ in Eq. (3)]. Another very important tuning exercise is the spread specification [determined by $\mathbf{D}$ in Eq. (3)]. The spread basically regulates the amplitude of the perturbation, which obviously influences the results strongly. Ideally the spread should be a function of time and space, based on the estimated targeted model error. Alternatively, the spread can be regulated to minimize negative effects of spurious correlations. In this case the spread should depend on the estimated strength of correlations with observed variables.

Throughout this work we have assumed a Gaussian dynamical model for z_0 . However, as z_0 relates lognormally to the surface fluxes, it might make sense to assume a skewed distribution as in Ruckstuhl and Janjić (2018). This would also alleviate the lower bound problem.

d. Resampling

To analyze the severity of spurious correlations and to aid the tuning, we propose to perform bootstrapping on the prior ensemble perturbations of z_0 (personal

communication with M. Verlaan). Note that the analysis increment of the ensemble mean of z_0 can be written as

$$\bar{z}_0^a - \bar{z}_0^f = \mathbf{X}_{z_0}^f \bar{\mathbf{w}}^a, \quad (5)$$

where $\mathbf{X}_{z_0}^f$ are the ensemble perturbations of z_0 on which $\bar{\mathbf{w}}^a \in \mathbb{R}^{N_{\text{ens}} \times N_{\text{ens}}}$ does not depend since z_0 is not represented in observation space (Hunt et al. 2007). Generating resampled analyses of $\bar{z}_0^a - \bar{z}_0^f$ is therefore relatively cheap. From the result we can learn about when and where spurious correlations are severe, which could provide hints on how to regulate the spread. Indeed, the detrimental effect of spurious correlations can be suppressed by lowering the spread.

e. Scale separation

Although the results are promising, we should critically consider Fig. 7. It is not clear if and when the standard deviation of z_0 converges. To verify that the parameter does not blow up, a longer test period is needed. Maintaining a flow dependent spread as suggested above should aid in keeping the parameter within certain bounds, but it might not be enough. In that case it might be favorable to split the parameter into its climatological part, which is slowly varying, and its remaining flow dependent part, which is rapidly varying (personal communication with M. Verlaan). The climatological part could be estimated with time averaged correlations, to smooth out spurious and flow dependent

correlations. The flow dependent deviations should be kept within a certain range of the climatological part either by resetting regularly, or imposing bound constraints. To investigate what works best, the parameter estimation should run over a long test period.

Acknowledgments. This study was carried out within Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” funded by the German Science Foundation (DFG) through subproject B6: “Parameter estimation using a data assimilation system for improved representation of clouds.” The authors wish to thank the DWD for use of their data and computing support within the framework of HERZ as well as Yuefei Zeng, Robert Redl, and Leonhard Scheck for usage of their data processing software. We are also grateful to the editor and the reviewers for their suggestions, which helped to improve the manuscript.

REFERENCES

- Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.*, **33**, L12801, <https://doi.org/10.1029/2006GL026186>.
- , D. C. Dowell, and C. Snyder, 2009: A multicas e comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, <https://doi.org/10.1175/2008MWR2691.1>.
- Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, <https://doi.org/10.1175/MWR-D-11-00013.1>.
- , 2016: Reducing correlation sampling error in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **144**, 913–925, <https://doi.org/10.1175/MWR-D-15-0052.1>.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, [https://doi.org/10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2).
- Annan, J., J. Hargreaves, N. Edwards, and R. Marsh, 2005: Parameter estimation in an intermediate complexity Earth system model using an ensemble Kalman filter. *Ocean Modell.*, **8**, 135–154, <https://doi.org/10.1016/j.ocemod.2003.12.004>.
- Bachmann, K., C. Keil, G. C. Craig, M. Weissmann, and C. Welzbacher, 2019: Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography, and synoptic weather regime. *Mon. Wea. Rev.*, **148**, 63–81, <https://doi.org/10.1175/MWR-D-19-0045.1>.
- Baldauf, M., A. Seifert, J. Foerster, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, <https://doi.org/10.1175/2010MWR3595.1>.
- Bick, T., and Coauthors, 2016: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 1490–1504, <https://doi.org/10.1002/qj.2751>.
- Buizza, R., M. Miller, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/QJ.49712556006>.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145, [https://doi.org/10.1175/1520-0493\(1995\)123<1128:OLEOEC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2).
- Desroziers, G., P. Brousseau, and B. Chapnik, 2005a: Use of randomization to diagnose the impact of observations on analyses and forecasts. *Quart. J. Roy. Meteor. Soc.*, **131**, 2821–2837, <https://doi.org/10.1256/qj.04.151>.
- , L. Berre, B. Chapnik, and P. Poli, 2005b: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- Doms, G., and Coauthors, 2011: A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization. Consortium for Smallscale Modeling (COSMO) Tech. Rep. LM F90 4.20, 154 pp., <http://www.cosmo-model.org/content/model/documentation/core/cosmoPhysParamtr.pdf>.
- Done, J. M., G. C. Craig, S. L. Gray, P. A. Clark, and M. E. B. Gray, 2006: Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quart. J. Roy. Meteor. Soc.*, **132**, 737–756, <https://doi.org/10.1256/qj.04.84>.
- Doron, M., P. Brasseur, J.-M. Brankart, S. N. Losa, and A. Melet, 2013: Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical–biogeochemical model. *J. Mar. Syst.*, **117**–**118**, 81–95, <https://doi.org/10.1016/j.jmarsys.2013.02.007>.
- Efron, B., and R. J. Tibshirani, 1994: *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 456 pp.
- Evensen, G., 2009: The ensemble Kalman filter for combined state and parameter estimation. *IEEE Contr. Syst. Mag.*, **29**, 83–104, <https://doi.org/10.1109/MCS.2009.932223>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Gelb, A., 1974: *Applied Optimal Estimation*. MIT Press, 382 pp.
- Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, <https://doi.org/10.1002/qj.3179>.
- Ha, S.-Y., and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1489–1508, <https://doi.org/10.1175/MWR-D-13-00108.1>.
- Hirt, M., S. Rasp, U. Blahak, and G. Craig, 2019: Stochastic parameterization of processes leading to convection initiation in kilometre-scale models. *Mon. Wea. Rev.*, **147**, 3917–3934, <https://doi.org/10.1175/MWR-D-19-0060.1>.
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, <https://doi.org/10.1016/j.physd.2006.11.008>.
- Janjić, T., L. Nerger, A. Albertella, J. Schroeter, and S. Skachko, 2011: On domain localization in ensemble based Kalman filter algorithms. *Mon. Wea. Rev.*, **139**, 2046–2060, <https://doi.org/10.1175/2011MWR3552.1>.

- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2011: "Variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation. *J. Geophys. Res.*, **116**, D09110, <https://doi.org/10.1029/2010JD014673>.
- Keil, C., F. Heinlein, and G. C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 480–490, <https://doi.org/10.1002/qj.2143>.
- Kober, K., and G. C. Craig, 2016: Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.*, **73**, 2893–2911, <https://doi.org/10.1175/JAS-D-15-0144.1>.
- Kondrashov, D., C. Sun, and M. Ghil, 2008: Data assimilation for a coupled ocean–atmosphere model. Part II: Parameter estimation. *Mon. Wea. Rev.*, **136**, 5062–5076, <https://doi.org/10.1175/2008MWR2544.1>.
- Kotsuki, S., K. Terasaki, H. Yashiro, H. Tomita, M. Satoh, and T. Miyoshi, 2018: Online model parameter estimation with ensemble data assimilation in the real global atmosphere: A case with the nonhydrostatic icosahedral atmospheric model (NICAM) and the global satellite mapping of precipitation data. *J. Geophys. Res. Atmos.*, **123**, 7375–7392, <https://doi.org/10.1029/2017JD028092>.
- Koyama, H., and M. Watanabe, 2010: Reducing forecast errors due to model imperfections using ensemble Kalman filtering. *Mon. Wea. Rev.*, **138**, 3316–3332, <https://doi.org/10.1175/2010MWR3067.1>.
- Lange, H., and T. Janjić, 2016: Assimilation of Mode-S EHS aircraft observations in COSMO-KENDA. *Mon. Wea. Rev.*, **144**, 1697–1711, <https://doi.org/10.1175/MWR-D-15-0112.1>.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).
- Murphy, J. M., D. M. Sexton, G. S. Barnett, D. N. Jones, M. J. Webb, M. Collins, and D. Sainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772, <https://doi.org/10.1038/nature02771>.
- Necker, T., M. Weissmann, R. Ruckstuhl, J. Anderson, and T. Miyoshi, 2020: Sampling error correction evaluated using a convective-scale 1000-member ensemble. *Mon. Wea. Rev.*, **148**, 1229–1249, <https://doi.org/10.1175/MWR-D-19-0154.1>.
- Owen, P. R., and W. R. Thomson, 1963: Heat transfer across rough surfaces. *J. Fluid Mech.*, **15**, 321–334, <https://doi.org/10.1017/S0022112063000288>.
- Piper, D., M. Kunz, F. Ehmele, S. Mohr, B. Mühr, A. Kron, and J. Daniell, 2016: Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany. Part I: Meteorological background. *Nat. Hazards Earth Syst. Sci.*, **16**, 2835–2850, <https://doi.org/10.5194/nhess-16-2835-2016>.
- Rabin, R. M., S. Stadler, P. J. Wetzels, D. J. Stensrud, and M. Gregory, 1990: Observed effects of landscape variability on convective clouds. *Bull. Amer. Meteor. Soc.*, **71**, 272–280, [https://doi.org/10.1175/1520-0477\(1990\)071<0272:OEOLVO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1990)071<0272:OEOLVO>2.0.CO;2).
- Raschendorfer, M., 2001: The new turbulence parameterization of LM. *COSMO Newsletter*, No. 1, Consortium for Small-Scale Modeling, Offenbach, Germany, 89–97, http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf.
- Rasp, S., T. Selz, and G. C. Craig, 2018: Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, **75**, 691–706, <https://doi.org/10.1175/JAS-D-17-0258.1>.
- Reinhardt, T., and A. Seifert, 2006: A three-category ice scheme for LMK. *COSMO Newsletter*, No. 6, Consortium for Small-Scale Modeling, Offenbach, Germany, 115–120, http://gop.meteo.uni-koeln.de/ag_crewell/lib/exe/fetch.php?media=projects:questdocuments:reinhardt_and_seifert_2006a_cnl.pdf.
- Rhodin, A., H. Lange, R. Potthast, and T. Janjić, 2013: Documentation of the DWD data assimilation system. Deutscher Wetterdienst (DWD), 448 pp.
- Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.*, **120**, 303–325, [https://doi.org/10.1175/1520-0493\(1992\)120<0303:ACRSFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2).
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Ruckstuhl, Y., and T. Janjić, 2018: Parameter and state estimation with ensemble Kalman filter based algorithms for convective-scale applications. *Quart. J. Roy. Meteor. Soc.*, **144**, 826–841, <https://doi.org/10.1002/qj.3257>.
- Ruiz, J., and M. Pulido, 2015: Parameter estimation using ensemble-based data assimilation in the presence of model error. *Mon. Wea. Rev.*, **143**, 1568–1582, <https://doi.org/10.1175/MWR-D-14-00017.1>.
- , —, and T. Miyoshi, 2013: Estimating model parameters with ensemble-based data assimilation: A review. *J. Meteor. Soc. Japan*, **91**, 79–99, <https://doi.org/10.2151/jmsj.2013-201>.
- Scheck, L., P. Frérebeau, R. Buras-Schnell, and B. Mayer, 2016: A fast radiative transfer method for the simulation of visible satellite imagery. *J. Quant. Spectrosc. Radiat. Transfer*, **175**, 54–67, <https://doi.org/10.1016/j.jqsrt.2016.02.008>.
- , M. Weissmann, and B. Mayer, 2018: Efficient methods to account for cloud-top inclination and cloud overlap in synthetic visible satellite images. *J. Atmos. Oceanic Technol.*, **35**, 665–685, <https://doi.org/10.1175/JTECH-D-17-0057.1>.
- Schirber, S., D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson, 2013: Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *J. Adv. Model. Earth Syst.*, **5**, 58–70, <https://doi.org/10.1029/2012MS000167>.
- Schraff, C., and R. Hess, 2012: A description of the nonhydrostatic regional COSMO-model. Part III: Data assimilation. COSMO Tech. Rep. V4_22, 99 pp., <http://www.cosmo-model.org/content/model/documentation/core/cosmoAssim.pdf>.
- , H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periañez, and R. Potthast, 2016: Kilometre-Scale Ensemble Data Assimilation for the COSMO model (KENDA). *Quart. J. Roy. Meteor. Soc.*, **142**, 1453–1472, <https://doi.org/10.1002/QJ.2748>.
- Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, <https://doi.org/10.1256/qj.04.106>.
- Simon, D., A. Samuelsen, L. Bertino, and S. Mouysset, 2015: Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the ensemble Kalman filter. *J. Mar. Syst.*, **152**, 1–17, <https://doi.org/10.1016/j.jmarsys.2015.07.004>.

- Sommer, M., and T. Janjić, 2018: A flexible additive inflation scheme for treating model error in ensemble Kalman filters. *Quart. J. Roy. Meteor. Soc.*, **144**, 2026–2037, <https://doi.org/10.1002/qj.3254>.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2).
- Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.
- Yang, R., and M. A. Friedl, 2003: Determination of roughness lengths for heat and momentum over boreal forests. *Bound.-Layer Meteor.*, **107**, 581–603, <https://doi.org/10.1023/A:1022880530523>.
- Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 251–262, <https://doi.org/10.1002/qj.353>.
- Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (icosahedral non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, <https://doi.org/10.1002/qj.2378>.
- Zeng, Y., U. Blahak, M. Neuper, and D. Jerger, 2014: Radar beam tracing methods based on atmospheric refractive index. *J. Atmos. Oceanic Technol.*, **31**, 2650–2670, <https://doi.org/10.1175/JTECH-D-13-00152.1>.
- , —, and D. Jerger, 2016: An efficient modular volume-scanning radar forward operator for NWP models: Description and coupling to the COSMO model. *Quart. J. Roy. Meteor. Soc.*, **142**, 3234–3256, <https://doi.org/10.1002/qj.2904>.
- , T. Janjić, A. de Lozar, U. Blahak, H. Reich, C. Keil, and A. Seifert, 2018: Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods, and combinations. *J. Adv. Model. Earth Syst.*, **10**, 2889–2911, <https://doi.org/10.1029/2018MS001375>.
- , —, M. Sommer, A. de Lozar, U. Blahak, and A. Seifert, 2019b: Representation of model error in convective-scale data assimilation: Additive noise based on model truncation error. *J. Adv. Model. Earth Syst.*, **11**, 752–770, <https://doi.org/10.1029/2018MS001546>.