Data Assimilation of nowcasted observations

Article

Published Version

Potthast, R. ORCID: https://orcid.org/0000-0001-6794-2500, Vobig, K., Blahak, U. and Simmer, C. (2022) Data Assimilation of nowcasted observations. Monthly Weather Review, 150 (5). pp. 969-980. ISSN 0027-0644 doi: https://doi.org/10.1175/MWR-D-21-0017.1 Available at https://centaur.reading.ac.uk/106221/

Data Assimilation of Nowcasted Observations

ROLAND POTTHAST,[a] KLAUS VOBIG,[a] ULRICH BLAHAK,[a] AND CLEMENS SIMMER[b]

[a] *Deutscher Wetterdienst, Data Assimilation Unit, Offenbach, Germany*
[b] *Institute for Geosciences, Section Meteorology, University of Bonn, Bonn, Germany*

(Manuscript received 23 February 2021, in final form 21 December 2021)

ABSTRACT: We investigate the *assimilation of nowcasted information* into a classical data assimilation cycle. As a reference setup, we employ the assimilation of standard observations such as direct observations of particular variables into a forecasting system. The pure advective movement extrapolation of observations as a simple *nowcasting* (NWC) is usually much better for the first minutes to hours, until outperformed by numerical weather prediction (NWP) based on data assimilation. Can nowcasted information be used in the data assimilation cycle? We study both an oscillator model and the Lorenz 63 model with assimilation based on the localized ensemble transform Kalman filter (LETKF). We investigate and provide a mathematical framework for the assimilation of nowcasted information, approximated as a local tendency, into the LETKF in each assimilation step. In particular, we derive and discuss adequate observation error and background uncertainty covariance matrices and interpret the assimilation of nowcasted information as assimilation with an H^1-type metric in observation space. Further, we show numerical results that prove that nowcasted information in data assimilation has the potential to significantly improve model based forecasting.

KEYWORDS: Nowcasting; Data assimilation; Model initialization; Nonlinear models

1. Introduction

Forecasting of dynamical systems has a long history in many application fields. In particular, weather and climate predictions (cf. Kalnay 2003) are based on sophisticated large-scale and high-dimensional numerical models run on supercomputers, to generate forecasts of the atmosphere, ocean, or Earth's system. The models are usually based on ordinary or partial differential equations and combine the Navier–Stokes equations with physical parameterization of radiation, turbulence, orographic influence, and microphysics.

These model system are initialized by the use of *data assimilation* techniques, such as variational data assimilation (3D-VAR or 4D-VAR) or ensemble data assimilation, e.g., the ensemble Kalman filter (EnKF) or, more recently, particle filters; for a detailed introduction we refer to Lorenc et al. (2000), Kalnay (2003), Evensen (2009), Anderson and Moore (2012), van Leeuwen et al. (2015), Reich and Cotter (2015), Kleist et al. (2009), Nakamura and Potthast (2015), Houtekamer and Zhang (2016), and Bannister (2017). The EnKF (Evensen 2009, 1994; Houtekamer and Mitchell 1998; Evensen and van Leeuwen 2000; Houtekamer and Mitchell 2001; Anderson 2001; Whitaker and Hamill 2002; Snyder and Zhang 2003; Houtekamer and Mitchell 2005; Houtekamer et al. 2005) employs an ensemble of states to dynamically estimate the background uncertainty covariance matrix and applies this matrix in each assimilation step.

Nowcasting (abbreviated NWC, e.g., Wilson and Edwin Kessler 1963; Bellon and Austin 1978; Hohti et al. 2000; Germann and Zawadzki 2002; Turner et al. 2004; Bowler et al. 2004; Haiden et al. 2011; Bechini and Chandrasekar 2017; Ryu et al. 2020) uses observations measured at high temporal and spatial density and assumes "Lagrangian persistence" of the observed phenomena. Observations are usually aggregated onto a 2D- or 3D-spatial grid to estimate a motion field from a time series of past observations, which explains the observed field changes in an "optimal" approximate way (e.g., Rinehart 1981; Anandan 1989; Li et al. 1995; Germann and Zawadzki 2002). In classical NWC, the most recent observations are extrapolated into the future by pure advective motion along the stream lines of such an estimated motion field, which itself is often assumed to be frozen in time. By doing so, observed and extrapolated time changes in the field are attributed solely to spatial movement. The forecast quality of this approximation for atmospheric parameters (e.g., precipitation from radar data, cloudy satellite observations) usually decreases rapidly with forecast lead time and increases with increasing spatial scale (e.g., Germann and Zawadzki 2002, 2004). Some NWC systems try to incorporate growth and decay into the extrapolation (e.g., Li et al. 1995; Golding 1998; Foresti et al. 2018; Sideris et al. 2020).

In the atmospheric sciences, traditionally nowcasting and numerical weather prediction have formed two different communities, which had their own methods, techniques, software environment, and language. More recently, an interplay between these techniques and communities started to evolve; we exemplarily point to the combined use of cloud ceiling height and visibility information with a rapid refresh numerical weather prediction system in Glahn et al. (2017) and to the SINFONY

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

Potthast's additional affiliation: Department of Mathematics, University of Reading, Whiteknights, Berkshire, United Kingdom.

Corresponding author: Roland Potthast, roland.potthast@dwd.de

DOI: 10.1175/MWR-D-21-0017.1

project of Deutscher Wettedienst,[1] which aims to generate seamless forecasts from minutes to days based on an integrated nowcasting and numerical weather prediction system, or the RealPEP research group funded by the German Research Foundation,[2] which brings together scientists from three communities to develop tools and methods which underpin and enable seamless forecasting based on multiple scales and diverse techniques.

Here, we develop a core part of such an integration by assimilating nowcasted information into a localized ensemble transform Kalman filter (LETKF). For this purpose, the nowcasting technique itself is approximated as the use of the local time tendency of an underlying observation. Of course, classical NWC may predict also nonlinear time series at fixed locations (consider a modal spatial structure moving over a certain location), but for very short lead times one may approximate any time series by a linear function.

Our setting and basic idea is sketched in Fig. 1. The dotted lines show an ensemble of forecasts, here taken as linear state evolutions. The black bullets visualize two observations, one at analysis time and another one shortly before. We can now detect the best ensemble member with respect to different metrics. Taking just one observation at analysis time leads to the best ensemble member shown by the green dashed line. Taking both observations, i.e., using classical 4D metric, leads to the ensemble member with the magenta dashed line. Taking the numerical derivative based on the two observations will lead to the blue dashed ensemble member as the best, because the linear curves coincide with the nowcasted information when linear extrapolation is used. Here, the blue curve provides the best fit to the red dashed line. For the displayed situation, where the truth is linear, the result is trivial. But it motivates that taking nowcasted observations into account in data assimilation has the potential to improve the forecast quality.

To be able to carry out a mathematical analysis and numerical investigation, we restrict ourselves to a simple oscillator model and to the popular Lorenz 63 system (Lorenz 1963, hereafter L63), see for example Evensen (1997), Verlaan and Heemink (2001), Vukicevic and Posselt (2008), Pu and Hacker (2009), Ambadan and Tang (2009), Carrassi and Vannitsem (2010), Lei and Bickel (2011), Hodyss (2011), Lei et al. (2012), Sakov et al. (2012), Yang et al. (2012), Zhang et al. (2012), Marzban (2013), and Goodliff et al. (2015) for its study and use.

When using nowcasted information on top of standard observations, we need to determine an adequate observation error covariance matrix. When observations are employed both for nowcasting and within the data assimilation cycle, observation errors will be correlated, and these correlations need to be taken into account to calculate the correct analysis increment.

In its simplest form nowcasting employs linear extrapolation calculated from differences and sums of observations,

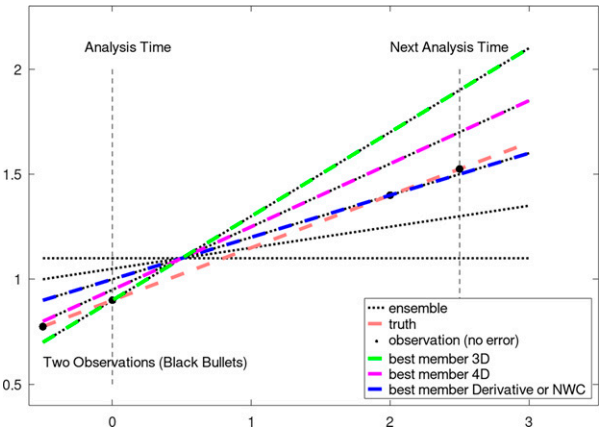

FIG. 1. The basic setup and idea of assimilating nowcasting information or derivative information shown by a display of an ensemble as dotted lines, some truth as red dashed line, and two observations at the analysis time and shortly before as black bullets. We also display best members with respect to different metrics: in dashed green is the best member with respect to the observation (black bullet) at analysis time (3D metric), in magenta is the best member when both observations at analysis time and shortly before the analysis time are used with classical 4D metric, and in blue is the best member when the numerical derivative or nowcasted information given the two observations is used (derivative or nowcasting case). Here, the best member with respect to the derivative or the nowcasting metric gives the best forecast, as can be seen by the distance to the two black bullets at the next analysis time.

and we can interpret these differences as functional of numerical derivative of the observation. Integrating derivatives into the data assimilation functional can be interpreted as moving from classical Euclidean norms $\|\cdot\|^2$ to Sobolev norms, such as

$$\|x\|^2_{H^1,\gamma} := \|x\|^2_{L^2} + \delta t^2 \left\| \frac{dx}{dt} \right\|^2_{L^2}, \tag{1}$$

with some factor δt, where t is the temporal variable of the dynamical system under consideration.

We will study the effect of the assimilation of nowcasted information practically both within an oscillator model and the well-known Lorenz 63 system, based on the LETKF. For both models, we will see that using this additional information in the cycled data assimilation system 1) has no effect of forecast skill when a linearly transformed observation error covariance matrix **R** is used or 2) can clearly improve forecast skill when a modified **R** matrix is employed.

The setup is described in section 2. The description of the LETKF for the assimilation of classical and nowcasted information is found in section 3. We carry out a basic mathematical analysis of assimilating nowcasted fields in section 4. Numerical results are presented in section 5 and conclusions in section 6.

2. An oscillator model, the Lorenz 63 model, and the observations setup

We employ a typical nonlinear model to investigate the assimilation of nowcasted information. The model must be

[1] See https://www.dwd.de/EN/research/researchprogramme/sinfony_iafe/sinfony_start_en.html.
[2] https://www2.meteo.uni-bonn.de/realpep/doku.php.

complex enough to capture typical nonlinear and potentially chaotic features of real-world NWP models, but also simple enough to gain insight into the behavior of the techniques. We have tested the ideas on two models, a simple oscillator model and the well-known Lorenz 63 model widely used for studying data assimilation algorithms and nonlinear deterministic chaos.

a. An oscillator model

We study the simple oscillator model defined for $x \in \mathbb{R}^2$:

$$\dot{x} = \mathbf{G}x, \tag{2}$$

with matrix $\mathbf{G}$ defined by

$$\mathbf{G} := \begin{pmatrix} 0 & \kappa_2 \\ -\kappa_1 & 0 \end{pmatrix}, \tag{3}$$

with wavenumbers κ_1 and κ_2. The system is given by

$$\dot{x}_1 = \kappa_2 x_2, \;\; \dot{x}_2 = -\kappa_1 x_1, \tag{4}$$

which by differentiation of x_1 with respect to time leads to

$$\ddot{x}_1 = -\kappa_1\kappa_2 x_1, \;\; \text{and} \;\; \ddot{x}_2 = -\kappa_1\kappa_2 x_2. \tag{5}$$

With $\kappa = \sqrt{\kappa_1\kappa_2}$, a solution of the equation is given by $x_1(t) = a\sin(\kappa t) + b\cos(\kappa t)$, $t \in \mathbb{R}$ and analogously for x_2. When for some $t \in \mathbb{R}$ values $x(t) = a_0$ and $\dot{x}(t) = a_1$ are provided, the solution is unique by the Picard–Lindelöf theorem Agarwal and O'Regan (2008), Nakamura and Potthast (2015).

b. The Lorenz 63 model

The well-known Lorenz 63 model Lorenz (1963) has been frequently employed in the data assimilation context over the past 50 years, see for example Evensen (1997), Verlaan and Heemink (2001), Vukicevic and Posselt (2008), Pu and Hacker (2009), Ambadan and Tang (2009), Carrassi and Vannitsem (2010), Lei and Bickel (2011), Hodyss (2011), Lei et al. (2012), Sakov et al. (2012), Yang et al. (2012), Zhang et al. (2012), Marzban (2013), Goodliff et al. (2015), Otkin et al. (2020). The model consists of the three coupled equations:

$$\frac{dx_1}{dt} = \sigma(x_2 - x_1), \tag{6}$$

$$\frac{dx_2}{dt} = \rho x_1 - x_2 - x_1 x_3, \tag{7}$$

$$\frac{dx_3}{dt} = x_1 x_2 - \beta x_3, \tag{8}$$

where $x_1(t)$, $x_2(t)$, and $x_3(t)$ are the dependent variables and σ, ρ, and β are parameters. For some parameter values, the system shows chaotic behavior because very small perturbations in the initial conditions can grow very rapidly into completely different solutions. The model was designed to simulate atmospheric dry cellular convection following the work of Saltzman (1962). The model simulates the evolution of a forced dissipative hydrodynamic system that possesses nonperiodic and unstable solutions. The variable x_1 measures the intensity of the convective motion, x_2 the temperature difference between the ascending and descending currents, and x_3 is a measure of the deviation of the vertical temperature profile from linearity. The model parameters are the Prandtl number (σ), a normalized Rayleigh number (ρ), and a nondimensional wavenumber (β). The Rayleigh number for the system is typically set to the slightly supercritical value of 28 following the work of Lorenz (1963), and we set σ and β to 10 and 8/3, respectively. The values guarantee a chaotic nature of the model. We employ a Runge–Kutta method of fourth-order to solve the forward integration following Nakamura and Potthast (2015).

c. The observational setup

We carry out a classical data assimilation cycle as in Kalnay (2003); Nakamura and Potthast (2015) for synchronizing model and truth based on observations. Observations $\mathbf{y}_k \in \mathbb{R}^m$ are given in an observation space $\mathbb{R}^m$ of dimension $m \in \mathbb{N}$ at times t_k for $k = 1, 2, 3, \ldots$. In a classical sequential data assimilation cycle an estimation (analysis) $x_k^{(a)}$ of the current system state $x^{(\text{true})}(t_k) \in \mathbb{R}^n$ with state space dimension n is calculated at times t_k based on the observations $\mathbf{y}_k$. Usually some model for the observation $\mathbf{y}_k$ at t_k is available, called *observation operator* $H : \mathbb{R}^n \to \mathbb{R}^m$, such that $H(x_k)$ simulates the observation $\mathbf{y}_k$ based on the state x_k.

Usually, observations are made at times $t \in \mathbb{R}$ which do not necessarily coincide with the analysis times t_k, $k = 1, 2, 3, \ldots$. In that case, a filter collects all observations, which belong to the interval between two consecutive analysis times:

$$I_k := [t_{k-1}, t_k], \tag{9}$$

into the observational vector $\mathbf{y}_k$ based on which the analysis $x_k^{(a)}$ for the analysis time t_k is computed.

Let us now add a simplified version of *nowcasting* to our setup. Nowcasting observes some variable or derived phenomenon at present t_k and in the recent past s_k given by

$$s_k = t_k - \delta_s, \;\; k = 1, 2, 3, \ldots, \tag{10}$$

where usually $\delta_s \ll |t_k - t_{k-1}|$ and then extrapolates the evolution of the variable or phenomenon for times $s \geq s_k$, including times $s > t_k$. We employ the following notation:

$$\mathbf{x}_{k,1} = x(s_k) \in \mathbb{R}^n, \quad \mathbf{x}_{k,2} = x(t_k) \in \mathbb{R}^n, \tag{11}$$

with time index $k = 1, 2, 3, \ldots$ and $\mathbf{x}_{k,\xi}^{(\text{true})}, \xi = 1, 2$, denoting the nature run. Here, we will assume that we observe

$$\begin{aligned} \mathbf{y}_{k,1}^o &:= H\left[\mathbf{x}_{k,1}^{(\text{true})}\right] + \epsilon, \\ \mathbf{y}_{k,2}^o &:= H\left[\mathbf{x}_{k,2}^{(\text{true})}\right] + \epsilon, \end{aligned} \tag{12}$$

with random observation error ϵ (perfect H assumed). If our observation is obtained from a dynamical system $x(t)$, observations $\mathbf{y}_{k,1}^o$ and $\mathbf{y}_{k,2}^o$ at times s_k and t_k can be used to approximate the temporal derivative:

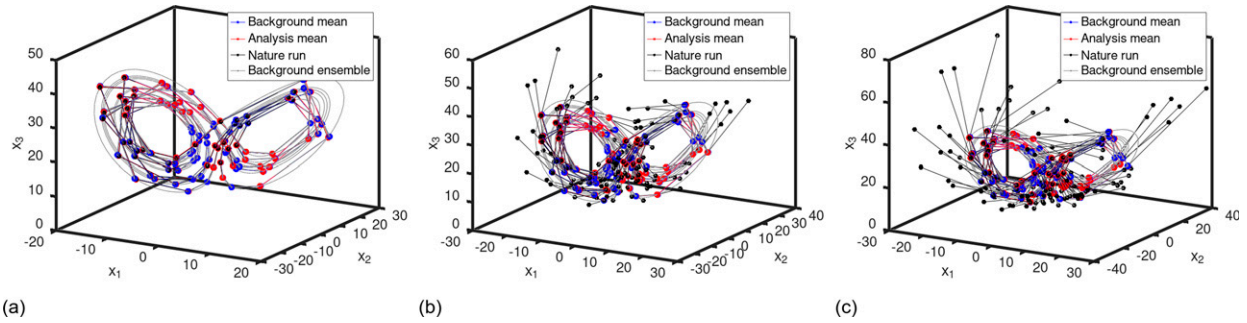

FIG. 2. Assimilation of nowcasted information with nowcasting lead times factors γ = (a) 0, (b) 3, and (c) 6 for the Lorenz 63 Model. This figure clarifies the setup and visualizes the relative size of the nowcasting linear vectors for the choices of γ used for the experiments. For our setup with observation operator $H = I$, the choice $\gamma = 3$ has been found to be optimal with respect to the first-guess error.

$$[H(x)]'(t) := \frac{dH[x(t)]}{dt} \tag{13}$$

at time $t \in [s_k, t_k]$. The approximation of the temporal derivative (13) by finite differences is given by

$$[H(x)]'(t) \approx \frac{\mathbf{y}^o_{k,2} - \mathbf{y}^o_{k,1}}{\delta_s}. \tag{14}$$

Clearly, approximating a derivative by finite differences is a well-known regularization of the ill-posed mapping from a function onto its derivative, compare Engl et al. (2000). Combining the assimilation of the observation and its numerical temporal derivative, this yields the following combination:

$$\mathbf{y}^o_k := \begin{pmatrix} \mathbf{y}^o_{k,2} \\ \dfrac{\mathbf{y}^o_{k,2} - \mathbf{y}^o_{k,1}}{\delta_s} \end{pmatrix}, \tag{15}$$

with $\delta_s = t_k - s_k$ for the standard approximation of the derivative. For our purpose, we approximate the classical nowcasting by its local linear time tendency at a given location, i.e., nowcasting is a simple *linear extrapolation in time:*

$$\mathbf{y}^{(\mathrm{nwc},s)}_k := c_1 \mathbf{y}^o_{k,1} + (s - s_k) \cdot \frac{\mathbf{y}^o_{k,2} - \mathbf{y}^o_{k,1}}{\delta_s}, \tag{16}$$

with a factor c_1 chosen to be $c_1 = 1$ for nowcasting and $c_1 = 0$ to include the assimilation of the numerical derivative described in (15) into one expression.

Collecting all temporal variables into the *weight or lead time* factor $\gamma = (s - s_k)/\delta_s$ or $\gamma = 1/\delta_s$, we obtain the following form:

$$\mathbf{y}^{(\mathrm{nwc})}_k(\gamma) := c_1 \mathbf{y}^o_{k,1} + \gamma \cdot \left(\mathbf{y}^o_{k,2} - \mathbf{y}^o_{k,1}\right), \tag{17}$$

for $s \geq s_k$ including times $s > t_k$, as indicated by the straight lines in Fig. 2. Here, we restrict our attention to this first approximation to more sophisticated nowcasting algorithms.

Combining classical observations $\mathbf{y}_{k,2}$ at t_k with nowcasted information $\mathbf{y}^{\mathrm{nwc}}_k(\gamma)$, we study the algebraic transformation given by (17) as a part of the observation operator, i.e., we consider observations given by

$$\mathbf{y}^o_k(\gamma) := \begin{pmatrix} \mathbf{y}^o_{k,2} \\ \mathbf{y}^{(\mathrm{nwc})}_k(\gamma) \end{pmatrix}, \quad k = 1, 2, 3, \ldots, \tag{18}$$

with *weight or lead-time factor* γ. Now Eq. (15) is a special case of (18) if $c_1 = 0$ is chosen.

3. Using the LETKF for derivative or NWC assimilation

We employ the localized ensemble transform Kalman filter (LETKF) as suggested by Hunt et al. (2007). This formulation allows to easily add new observations to the assimilation without changing the core implementation. The formulation Hunt et al. (2007), see also Nakamura and Potthast (2015, chapter 5) solves the Kalman filter equations in ensemble space spanned by the ensemble members $x^{(b,\ell)}$ for $\ell = 1, \ldots, L$ with the ensemble mean given by

$$\overline{x}^{(b)} := \frac{1}{L} \sum_{\ell=1}^{L} x^{(b,\ell)}. \tag{19}$$

We use the following notation:

$$\mathbf{X}^b := \left[x^{(b,1)} - \overline{x}^{(b)}, \ldots, x^{(b,L)} - \overline{x}^{(b)}\right], \tag{20}$$

for the matrix of ensemble member differences from the mean and (for the case of a linear H)

$$\mathbf{Y}^b := H\mathbf{X}^b, \tag{21}$$

for the ensemble differences in observation space, $\mathbf{y}^o$ for the observation vector and $\overline{\mathbf{y}}^b$ for the mean of observations simulated from the ensemble. The observation error covariance matrix is denoted by $\mathbf{R}$. Now, we employ Eqs. (20) and (21) of Hunt et al. (2007):

$$\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y}^o - \overline{\mathbf{y}}^b\right), \tag{22}$$

for calculating the linear coefficients of the analysis mean with $\tilde{\mathbf{P}}^a$ given by

$$\tilde{\mathbf{P}}^a = \left[(L-1)\mathbf{I} + (\mathbf{Y}^b)^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}^b\right]^{-1}, \tag{23}$$

with $\mathbf{I}$ for the identity matrix. $\tilde{\mathbf{P}}^a$ denotes the $L \times L$ analysis covariance in $\mathbb{R}^L$. Equation (2) in model space leads to (22) and (23) of Hunt et al. (2007):

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{X}^b \overline{w}^a, \tag{24}$$

$$\mathbf{P}^a = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{X}^b)^{\mathrm{T}}, \tag{25}$$

where $\overline{\mathbf{x}}^a$ is the analysis mean and $\mathbf{P}^a$ the analysis covariance matrix. With $\mathbf{W}$ calculated by

$$\mathbf{W} = \left[(L-1)\tilde{\mathbf{P}}^a\right]^{1/2}, \tag{26}$$

as in (24) of Hunt et al. (2007) the analysis ensemble is given by

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}, \tag{27}$$

where the power 1/2 denotes the symmetric square root of the symmetric matrix $\tilde{\mathbf{P}}^a$.

For assimilating nowcasted information, let us employ the 4D extension of the above formulation. Here, the observation vector $\mathbf{y}^b$ is not directly calculated by an application of H to the states $x^{(b,\ell)}$. For observations at time $s \in [t_{k-1}, t_k]$, the simulated observation $y^{(b,\ell)}(s)$ is calculated during the model run $x^{(b,\ell)}(t)$. LETKF calculates a transform matrix based on the matrix $\mathbf{Y}^b$ (model equivalents of ensemble members minus their mean) and on observations $\mathbf{y}^o$ at one or several points in time. The transform matrix can be applied to the ensemble at any point in time, such that 3D and 4D just refer to the choice of observations and their temporal information within $\mathbf{Y}^b$.

With the (simulated) observation $\mathbf{y}^o_{k,1}$ at time s_k and $\mathbf{y}^o_{k,2}$ at time t_k, as described in Eq. (12), a nowcasted observation is given by (17). Similarly, the simulated observation for ensemble member ℓ is denoted by $\mathbf{y}^{(b,\ell)}_{k,1}$ for time s_k and $\mathbf{y}^{(b,\ell)}_{k,2}$ for time t_k, such that the nowcasted components of the simulation are given by

$$\mathbf{y}^{\mathrm{nwc},\ell}_k(\gamma) := c_1 \mathbf{y}^{(b,\ell)}_{k,1} + \gamma\left[\mathbf{y}^{(b,\ell)}_{k,2} - \mathbf{y}^{(b,\ell)}_{k,1}\right]. \tag{28}$$

With the mean $\overline{\mathbf{y}^b_{k,\xi}}$ of the set:

$$\left\{\mathbf{y}^{(b,\ell)}_{k,\xi} : \ \ell = 1, \ldots, L\right\}, \quad \xi = 1, 2,$$

for the observation at s_k ($\xi = 1$) and at t_k ($\xi = 2$), the matrix of differences $\mathbf{Y}^b_{k,\xi}$ is given by

$$\mathbf{Y}^b_{k,\xi} := \left[\mathbf{y}^{(b,1)}_{k,\xi} - \overline{\mathbf{y}^b_{k,\xi}}, \ldots, \mathbf{y}^{(b,L)}_{k,\xi} - \overline{\mathbf{y}^b_{k,\xi}}\right], \quad \xi = 1, 2, \tag{29}$$

and the nowcasted differences matrix is

$$\mathbf{Y}^{\mathrm{nwc}}_k(\gamma) := c_1 \mathbf{Y}^b_{k,1} + \gamma \cdot \left(\mathbf{Y}^b_{k,2} - \mathbf{Y}^b_{k,1}\right). \tag{30}$$

To combine classical and nowcasted information following (18), we need to work with the matrix $\mathbf{Y}^b_k$ of differences between simulated observations and their ensemble mean:

$$\mathbf{Y}^b_k(\gamma) := \begin{pmatrix} \mathbf{Y}^b_{k,2} \\ \mathbf{Y}^{\mathrm{nwc}}_k(\gamma) \end{pmatrix}. \tag{31}$$

Note that for $\gamma = 0$ the lower part of the matrices corresponds to a measurement at s_k. For $\gamma > 0$ it is a linear combination of measurements at s_k and t_k. The case $\gamma = c_1$ is a singular case, where the observation $\mathbf{y}^o_{k,2}$ at time t_k is used twice.

4. Theory for assimilating nowcasted observations

This section develops the mathematical theory for the assimilation of nowcasted observations as described by Eq. (18). Later we will use the LETKF to assimilate both classical and nowcasted observations within the data assimilation cycle to demonstrate its benefit.

For the assimilation the background error covariance matrix and the observation error covariance matrix are required. In section 4a we calculate the observation error covariance matrix when using linear transformations of observations—as given by temporal derivatives or nowcasting in the sense of Eq. (17). We will discuss the theory of assimilating such linearly transformed observations in section 4b. In section 4c we show that assimilating nowcasted information with a transformed $\mathbf{R}$ matrix is equivalent to assimilating the observations directly. The same holds for assimilating the temporal derivative with a corresponding $\mathbf{R}$ matrix.

a. Estimating the observation error covariance matrix

We note that basically all data assimilation methods (Lorenc et al. 2000; Kalnay 2003; Evensen 2009; Anderson and Moore 2012; van Leeuwen et al. 2015; Reich and Cotter 2015; Kleist et al. 2009; Nakamura and Potthast 2015; Houtekamer and Zhang 2016; Bannister 2017) somehow minimize a cost functional of the following form:

$$J(x) := \left\|x - x^{(b)}_k\right\|^2_{\mathbf{B}^{-1}} + \left\|y_k - H(x)\right\|^2_{\mathbf{R}^{-1}}, \tag{32}$$

with

$$\|\tilde{x}\|^2_{\mathbf{B}^{-1}} := \tilde{x}^{\mathrm{T}} \mathbf{B}^{-1} \tilde{x}, \ \ \|\tilde{y}\|^2_{\mathbf{R}^{-1}} := \tilde{y}^{\mathrm{T}} \mathbf{R}^{-1} \tilde{y}, \tag{33}$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is the background error covariance matrix and $\mathbf{R} \in \mathbb{R}^{m \times m}$ is the observation error covariance matrix for the observations $\mathbf{y}_k$ at time t_k. For the LETKF (Hunt et al. 2007), the background error covariance matrix $\mathbf{B}$ is implicitly estimated when the update formulas in ensemble space are carried out. The observation error covariance matrix $\mathbf{R}$ corresponds to the setup of the forward operator H which is applied to all ensemble members at the observation time. Now we need to calculate the observation error covariance matrix for the observation vector of classical plus nowcasted information, i.e., for the observation $\mathbf{y}_k$ defined by (18). Let us assume that

A1: $\mathbf{R}_0 \in \mathbb{R}^{m \times m}$ is the observation error covariance matrix for the observation $y(s)$ at time s, and that

A2: the observations at different times $s_1 \neq s_2$ are independent, i.e., the cross covariances between $y(s_1)$ and $y(s_2)$ are zero.

Theorem 4.1 *Under the assumptions A1 and A2, the observation error covariance* $\mathbf{R}$ *for the observation* $\mathbf{y}_k$ *defined in (18) with* γ *defined in (17) is given by*

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} = \begin{Bmatrix} \mathbf{R}_0 & \gamma\mathbf{R}_0 \\ \gamma\mathbf{R}_0 & [(c_1-\gamma)^2+\gamma^2]\mathbf{R}_0 \end{Bmatrix}$$

$$=: \begin{pmatrix} 1 & \gamma \\ \gamma & (c_1-\gamma)^2+\gamma^2 \end{pmatrix} \otimes \mathbf{R}_0 . \tag{34}$$

We now assume $\gamma \neq c_1$. Then, if $\mathbf{R}_0$ is invertible, so is $\mathbf{R}$, with the inverse given by

$$\frac{1}{c_1-\gamma^2}\begin{bmatrix} (c_1-\gamma)^2+\gamma^2 & -\gamma \\ -\gamma & 1 \end{bmatrix} \otimes \mathbf{R}_0^{-1}. \tag{35}$$

Proof. We first note that by definition we have

$$\mathbf{R}_{ij} = \mathbb{E}\Big\{\Big(\mathbf{y}_{k,i} - \overline{\mathbf{y}_{k,i}}\Big)\Big(\mathbf{y}_{k,j} - \overline{\mathbf{y}_{k,j}}\Big)\Big\}, \tag{36}$$

where i, j indicate the components of the vector $\mathbf{y}_k$. The second part of $\mathbf{y}_k$ consists of linear combinations of $\mathbf{y}_1$ and $\mathbf{y}_2$, which by assumption both have the observation error covariance matrix $\mathbf{R}_0$. The first part of $\mathbf{y}_k$ consists of $\mathbf{y}_{k,2}$ with the observation error covariance matrix $\mathbf{R}_0$. We first note that the observation error covariance matrix of $c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1})$ is calculated by

$$\begin{aligned} \mathbf{R}_{22} &:= \mathbb{E}\Big\{c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1}) - \overline{\big[c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1})\big]} \\ &\quad \cdot c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1}) - \overline{\big[c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1})\big]}^{\mathrm{T}}\Big\} \\ &= \mathrm{Var}\big[(c_1-\gamma)\mathbf{y}_{k,1} + \gamma\mathbf{y}_{k,2}\big] \\ &= \big[(c_1-\gamma)^2+\gamma^2\big]\mathbf{R}_0. \end{aligned} \tag{37}$$

In a similar way, we calculate the submatrices:

$$\begin{aligned} \mathbf{R}_{12} &:= \mathbb{E}\Big(\big(\mathbf{y}_{k,2} - \overline{\mathbf{y}_{k,2}}\big)\Big\{\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1}) \\ &\quad - \overline{\big[c_1\mathbf{y}_{k,1} + \gamma(\mathbf{y}_{k,2} - \mathbf{y}_{k,1})\big]}\Big\}^{\mathrm{T}}\Big) \\ &= \mathbb{E}\Big\{\big(\mathbf{y}_{k,2} - \overline{\mathbf{y}_{k,2}}\big)\big[\gamma(c_1\mathbf{y}_{k,2} - \overline{\mathbf{y}_{k,2}})\big]^{\mathrm{T}}\Big\} \\ &= \gamma\mathbf{R}_0 \end{aligned} \tag{38}$$

and the same for $\mathbf{R}_{21}$. The Eqs. (37) and (38) together with the observation error covariance matrix for

$$\begin{aligned} \mathbf{R}_{11} &= \mathbb{E}\Big[\big(\mathbf{y}_{k,2} - \overline{\mathbf{y}_{k,2}}\big)\big(\mathbf{y}_{k,2} - \overline{\mathbf{y}_{k,2}}\big)^T\Big] \\ &= \mathbf{R}_0 \end{aligned} \tag{39}$$

leads to the result (34). Finally, we note that for $\gamma \neq 1$ the determinant of the matrix

$$\mathbf{C} := \begin{bmatrix} 1 & \gamma \\ \gamma & (c_1-\gamma)^2+\gamma^2 \end{bmatrix} \tag{40}$$

is given by

$$\det(\mathbf{C}) = (c_1-\gamma)^2 > 0,$$

such that $\mathbf{C}$ has the inverse

$$\mathbf{C}^{-1} = \frac{1}{(c_1-\gamma)^2}\begin{bmatrix} (c_1-\gamma)^2+\gamma^2 & -\gamma \\ -\gamma & 1 \end{bmatrix}$$

and $\mathbf{C} \otimes \mathbf{R}_0$ is invertible if and only if $\mathbf{R}_0$ is invertible and the inverse is given by $\mathbf{C}^{-1} \otimes \mathbf{R}_0^{-1}$ as verified by elementary matrix operations.

b. Equivalence of assimilating linearly transformed data with their direct assimilation

For the ensemble Kalman filter, following Nakamura and Potthast (2015), chapter 5, the mean analysis update is based on the following equation:

$$x^a = x^b + \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}\big[y - H(x^b)\big], \tag{41}$$

and the update of the $\mathbf{B}$ matrix given by

$$\mathbf{B}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}^b, \quad \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}. \tag{42}$$

We include the case where the observation operator operates on several times. Then x^b (and x^a) might contain two (or more) time steps and $\mathbf{B}$ includes cross-temporal correlations. However, the analysis state is usually calculated at one analysis time only. The standard LETKF formulation of Hunt et al. (2007) can be seen as a transformed version of the Eqs. (41) and (42). For the case of $\mathbf{H}$ linear, let us now assume that we carry out a linear transform $\mathbf{A}$ of the observations:

$$\tilde{\mathbf{H}} = \mathbf{A}\mathbf{H}, \quad \tilde{y} = \mathbf{A}y. \tag{43}$$

The corresponding transform of the observation error covariance matrix $\mathbf{R}$ is given by

$$\tilde{\mathbf{R}} = \mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}}, \tag{44}$$

and the Kalman matrix for the transformed observations is given by

$$\tilde{\mathbf{K}} = \mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\Big(\tilde{\mathbf{R}} + \tilde{\mathbf{H}}\mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\Big)^{-1}. \tag{45}$$

We now study the transform of the Kalman filter equations, Eqs. (41) and (42), based on the transform given by (43), (44), and (45). We calculate the Kalman equations for the transformed quantities and show that they are identical to the original equations. We calculate

$$\begin{aligned} x^a &= x^b + \mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\Big(\tilde{\mathbf{R}} + \tilde{\mathbf{H}}\mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\Big)^{-1}\Big[\tilde{y} - \tilde{\mathbf{H}}\Big(x^b\Big)\Big] \\ &= x^b + \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\Big(\mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\Big)^{-1}\mathbf{A}\Big[y - \mathbf{H}\Big(x^b\Big)\Big] \\ &= x^b + \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\Big(\mathbf{A}^{\mathrm{T}}\Big)^{-1}\Big(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\Big)^{-1}\mathbf{A}^{-1}\mathbf{A}\Big[y - \mathbf{H}\Big(x^b\Big)\Big] \\ &= x^b + \mathbf{B}\mathbf{H}^{\mathrm{T}}\Big(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\Big)^{-1}\Big[y - \mathbf{H}\Big(x^b\Big)\Big] \end{aligned} \tag{46}$$

and for the background error covariance matrix $\mathbf{B}^a$ we calculate

$$\begin{aligned}\mathbf{B}^a &= \left(\mathbf{I} - \tilde{\mathbf{K}}\tilde{\mathbf{H}}\right)\mathbf{B}^b \\ &= \left[\mathbf{I} - \mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\left(\tilde{\mathbf{R}} + \tilde{\mathbf{H}}\mathbf{B}\tilde{\mathbf{H}}^{\mathrm{T}}\right)^{-1}\tilde{\mathbf{H}}\right]\mathbf{B}^b \\ &= \left[\mathbf{I} - \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{A}\mathbf{H}\right]\mathbf{B}^b \\ &= \left[\mathbf{I} - \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}}\right)^{-1}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{A}^{-1}\mathbf{A}\mathbf{H}\right]\mathbf{B}^b \\ &= \left[\mathbf{I} - \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\right]\mathbf{B}^b. \end{aligned} \tag{47}$$

We can now state the following basic lemma.

Lemma 4.2 *Let us carry out an assimilation step with a Kalman filter (41)–(42) for transformed observations $\tilde{y} = \mathbf{A}y$ with some linear invertible transform $\mathbf{A}$. For the assimilation of the transformed observation we employ the transformed Kalman matrix $\tilde{\mathbf{K}}$ with the transformed observation error covariance matrix $\tilde{\mathbf{R}}$. Then, the analysis mean x^a and covariance matrix $\mathbf{B}^a$ are identical to the result when assimilating the original observations y based on the Kalman matrix $\mathbf{K}$ and the observation error covariance matrix $\mathbf{R}$.*

Proof. We note that Eqs. (41) and (42) are purely algebraic and that they are applicable to any vector $\mathbf{x}^b$ which may consist of a finite number of time slices, with covariance $\mathbf{B}$ taking these into account. The transform with $\mathbf{A}$ takes place in observation space only. The statement of the theorem is a direct result of the derivation given in (46) and (47).

For the ensemble Kalman filter (Hunt et al. (2007)) we note that $\mathbf{B}$ is estimated by $\mathbf{B} = 1/(L-1)\mathbf{X}\mathbf{X}^{\mathrm{T}}$ with the matrix of $L \in \mathbb{N}$ ensemble members:

$$\mathbf{X} = \left[x^{(1)} - \bar{x}, \ldots, x^{(L)} - \bar{x}\right], \bar{x} = \frac{1}{L}\sum_{\ell=1}^{L} x^{(\ell)}.$$

With $\mathbf{Y} = \mathbf{H}\mathbf{X}$ the term $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} = 1/(L-1)\mathbf{Y}\mathbf{Y}^{\mathrm{T}}$ employs simulated observations only. For the LETKF it is used to calculate a transform matrix to compose the analysis ensemble obeying (46) and (47) (see Potthast and Welzbacher (2018), section 2, for details).

c. Nowcasted information or temporal derivatives as linear transform

Let us now study the assimilation of nowcasted information or a numerical temporal derivative. According to theorem 4.1, the transformed observation error covariance matrix for (15) and (18) is given by (34). With the transform matrix A defined by

$$\mathbf{A} := \begin{bmatrix} 0 & 1 \\ (c_1 - \gamma) & \gamma \end{bmatrix}, \tag{48}$$

we obtain

$$\mathbf{y}_k^o(\gamma) = \mathbf{A}\begin{pmatrix} \mathbf{y}_{k,1}^o \\ \mathbf{y}_{k,2}^o \end{pmatrix}, \mathbf{R}(\gamma) = \mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}}. \tag{49}$$

We are now prepared to formulate an equivalence result for assimilating nowcasted information.

Theorem 4.3 *We consider the assimilation of the combination (18) of some observation $\mathbf{y}_{k,2}$ at time t_k with the numerical temporal derivative ($c_1 = 0$) or the nowcasted information ($c_1 = 1$) in the form $\mathbf{y}^{\mathrm{nwc}}(\gamma)$based on observations at times s_k and t_k, $k = 1, 2, 3, \ldots$, into an ensemble Kalman filter. We assume that we employ the transformed observation error covariance matrix $\mathbf{R}(\gamma)$ based on the observation error covariance $\mathbf{R}_0$ for the measurement $\mathbf{y}_{k,1}^o$ and $\mathbf{y}_{k,2}^o$, which have uncorrelated observation errors. Then, for all $\gamma \geq 0$, $\gamma \neq c_1$, the assimilation result is identical to the assimilation of the original observations $\mathbf{y}_{k,1}^o$ and $\mathbf{y}_{k,2}^o$, $k = 1, 2, 3, \ldots$ with observation error covariance matrix:*

$$\mathbf{R} := \begin{pmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_0 \end{pmatrix}. \tag{50}$$

Proof. The theorem is a result of the combination of theorem 4.1 and lemma 4.2. We have calculated the covariance matrix for observation setup (18) in theorem 4.1. It can be obtained in the form of a linear transform (49) with A defined in (48). As worked out in Potthast and Welzbacher (2018), section 2, the ensemble Kalman filter formulation can be written in the form (41) and (42), such that lemma 4.2 is applicable in this case. Thus, lemma 4.2 tells us that the outcome of each assimilation step is independent of the linear transform. This then clearly applies to the whole assimilation cycle.

We will later carry out numerical experiments assimilating (18) into the oscillator model system with different nowcasting lead time factors γ. The above theorem is confirmed by the curve shown in Fig. 3 for $c_1 = 1$. For the choice $c_1 = 0$ assimilating the numerical temporal derivative with different weights to the derivative, the result is numerically confirmed with a graph identical to the one shown in Fig. 3. The same result is observed with the Lorenz 63 model.

With the above equivalence result we have shown that with the transformed observation error covariance matrix $\mathbf{R}(\gamma)$ given by (34) there is no benefit in using nowcasted information. However, we will see in section 5 that the benefit can be achieved if a different observation error covariance matrix $\mathbf{R}$ is used such that nowcasted information can indeed help to improve the forecast scores.

5. Numerical study assimilating nowcasted observations

In this section, we show numerical results when assimilating nowcasted information for the oscillator model in section 5a and for the Lorenz 63 model in section 5b. In particular, we will study the choices of $\mathbf{R}$ given by (34) and (50) for assimilating $\mathbf{y}_{k,2}^o$ only, $\mathbf{y}_{k,1}^o$ and $\mathbf{y}_{k,2}^o$, or the combination $\mathbf{y}_{k,2}^o$ with $\mathbf{y}_k^{(\mathrm{nwc},\gamma)}$.

Our general goal is to study the assimilation of nowcasted information and compare it to classical assimilation of observations, both with observations at analysis time only (3D-LETKF) or with an additional observation from the interval between analysis times (4D-LETKF). Further, we will study the dependence of the results on the weight or lead time factor γ, i.e., on the nowcasting time interval.

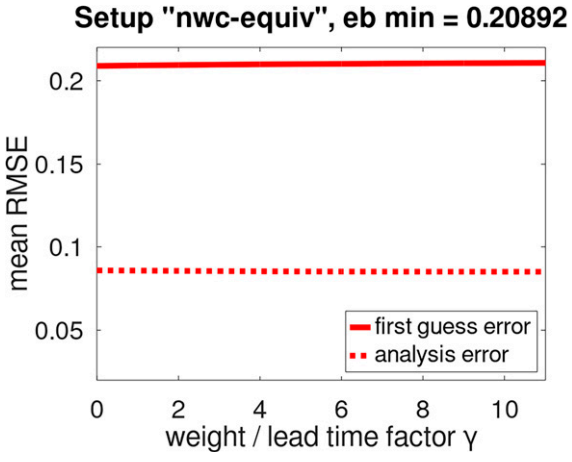

FIG. 3. For the oscillator model, we show the first guess error (full line) and analysis error (dotted line) for a data assimilation cycle with 100 steps for data given by (18). Basically the same image is obtained from (15), not duplicated here. We show the error in dependence on the lead time of the nowcasting or the scaling of the derivative. Clearly, for each γ there are different nowcasted (linearly transformed) observations and corresponding observation error covariance matrix $\mathbf{R}(\gamma)$ entering the LETKF. Up to numerical precision, the result of theorem 4.3 is confirmed, i.e., assimilating linearly transformed data with the corresponding observation error covariance matrix does not change the result.

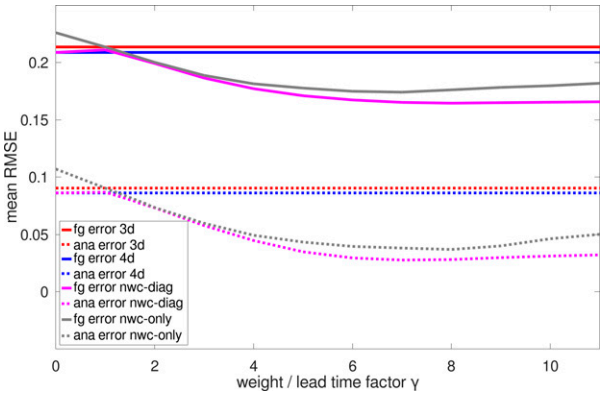

FIG. 4. We display the first guess (full line) and analysis error (dotted line) for different setups of the assimilation cycle for the linear oscillator. The red lines show the results of a 3D setup, the errors are independent of γ, the average first guess error is $e^b = 0.213$, and average analysis error $e^a = 0.09$ for 100 assimilation cycles. The blue lines display the results for two observations assimilated by 4D-LETKF, independent of γ as well. The magenta lines show results for a fixed diagonal observation error covariance matrix in (50). The results show the dependence of the mean analysis and first guess error on the nowcasting lead time factor γ in (18). The gray lines display the errors for assimilating nowcasted fields only with fixed observation error covariance matrix $\mathbf{R}_0$.

a. Numerical results for the oscillator model

We have implemented a fourth-order Runge–Kutta method for the oscillator model (2) and a localized ensemble transform Kalman filter as described in Hunt et al. (2007). We have chosen a nature run with $\kappa = 1.2$ and used $\kappa = 1 + 0.05 \cdot \mathrm{randn}(\ell)$, $\ell = 1, \ldots, L$, i.e., a random choice with mean $\kappa = 1$ and variance 0.05 for the ensemble members in the assimilation cycle. The noise for observations is set to $\epsilon = 0.013$ times Gaussian random noise. We chose $T = 1$ for the temporal intervals $T = t_k - t_{k-1}$ of the assimilation cycle and an initial state given by $x_0 = (0, 1)^T$. The observation operator H is chosen to be $H = (1, 0)$, i.e., we observe the first component of the two-dimensional oscillating system.

The results of different setups are shown in Fig. 4. With the time T between two assimilation times t_{k-1} and t_k, the choice $\delta_s = T/6$ and the constant $\gamma = \delta t/\delta_s$ the value $\gamma = 12$ corresponds to $\delta t = 2T$, i.e., we plot the scores for nowcasting lead times in the interval $[0,2T)$.

The results for choosing R according to theorem 4.3 lead to the results shown in Fig. 3, where the quality of the combined assimilation of $\mathbf{y}^o_{k,2}$ and $\mathbf{y}^{\mathrm{nwc}}_k(\gamma)$ is identical to the assimilation of $\mathbf{y}^o_{k,2}$ and $\mathbf{y}^o_{k,1}$, i.e., to the 4D case, for all lead-time factors γ.

However, using a different observation error covariance matrix $\mathbf{R}$ can improve the results. The experiments shown in Fig. 4 have been carried out with a diagonal $\mathbf{R}$ matrix $\mathbf{R} = \mathbf{R}_0 = (0.013)^2$ (with the numerical error used to simulate the observations) if only the observation $\mathbf{y}^o_{k,2}$ is assimilated or

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_0 \end{pmatrix}$$

in the case of two observations $\mathbf{y}^o_{k,2}$ and $\mathbf{y}^{\mathrm{nwc}}(\gamma)$.

The red lines in Fig. 4 show the result when only the observation $\mathbf{y}^o_{k,2}$ [see Eq. (12)] is assimilated at each analysis time t_k, $k = 1, \ldots, N_{\mathrm{nat}}$, which we refer to as 3D case. The results are independent of the parameter γ, since it does not appear in the formula when $\mathbf{y}^o_{k,2}$ is assimilated. With the parameter choices above for an analysis cycle with $N_{\mathrm{nat}} = 100$ analysis steps we found an average first guess error $e_b = 0.2137$ and an average analysis error of $e_a = 0.0902$.

With the blue lines in Fig. 4 we now take an additional earlier observation $\mathbf{y}^o_{k,1}$ into account for each of the $N_{\mathrm{nat}} = 100$ analysis steps, located at time $s_k = t_k - \delta_s$ for $k = 1, \ldots, N_{\mathrm{nat}}$. We feed both observations into the LETKF, where now the observation vector at each analysis time t_k consists of

$$\mathbf{y}^o = \begin{pmatrix} \mathbf{y}^o_{k,2} \\ \mathbf{y}^o_{k,1} \end{pmatrix}. \tag{51}$$

We note that this is identical to (18) with γ chosen to be zero, since then

$$\mathbf{y}^{\mathrm{nwc}}_k(0) = \mathbf{y}_{k,1} + 0 \cdot (\mathbf{y}_{k,2} - \mathbf{y}_{k,1}) = \mathbf{y}_{k,1}\,.$$

The result is again independent of γ, and we observe an error of $e_b = 0.208\,92$ and an analysis error of $e_a = 0.086\,012$. This means this 4D version of the LETKF is slightly better than the 3D version for the above setup.

We now assimilate $\mathbf{y}^o_{k,2}$ and $\mathbf{y}^{\text{nwc}}_k(\gamma)$ for different nowcasting lead time factors γ into the LETKF, displayed by the magenta lines in Fig. 4. This means that choosing fixed γ, with $N_{\text{nat}} = 100$ assimilation cycles for $k = 1, \ldots, N_{\text{nat}}$ at each analysis time t_k we assimilate both the classical observation at t_k and the nowcasted observation with nowcasting lead-time factor γ. Thus, the analysis x^a_k and the first guess x^b_{k+1} now depends on γ. We run the experiments for twelve choices of $\gamma = 0, 1, 2, \ldots, 11$ and display the first-guess errors $e_b(\gamma)$ as well as the analysis errors $e_a(\gamma)$. The minimal first-guess error is given by $\min e_b = 0.16453$, which is about 20% below the error observed for the classical 3D and 4D setup of the assimilation cycle. It can also be observed that the error decays strongly until approximately $\gamma = 6$, which roughly corresponds to the first guess lead time $T = 6\delta_s$ in the assimilation cycle, where our first-guess error is calculated as $e^b_{k+1} := |x^b_{k+1} - x^{\text{true}}_{k+1}|$. This means we get best results when the lead time of the nowcasting is roughly identical to the time when we evaluate the short-range forecast at the following analysis time.

A final setup is displayed by the gray lines in Fig. 4, where we now use nowcasted information $\mathbf{y}^{\text{nwc}}_k(\gamma)$ only, not assimilating the observation $\mathbf{y}^o_{k,2}$ directly. With the identical parameter choices as in Figs. 2a–c, we obtain a minimal first-guess error $e_b = 0.17415$, which is worse than the joint assimilation of $\mathbf{y}^o_{k,2}$ and $\mathbf{y}^{\text{ncw}}_k(\gamma)$, but still better than the use of the observations in the 3D or 4D setup without any nowcasting.

b. Numerical results for Lorenz 63 model

For further numerical testing, we employ the Lorenz model (Lorenz 1963) as described in section 2b. Data assimilation for the Lorenz 63 model is explained in detail including typical MATLAB or Octave codes in Nakamura and Potthast (2015, chapter 6).

Figure 2 shows the setup for our numerical testing. We run a data assimilation cycle with cylcing interval of $\delta t = 0.12$. We first investigate the case where the observation operator H is given by $H = I$, i.e., we observe the full state in each assimilation time step. We calculate a nature run with parameters $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ for which the system exhibits chaotic behavior. We then simulate observations with Gaussian errors based on a random draw with $\epsilon = 0.02$ and corresponding error covariance matrix given by $\mathbf{R}_0 = \epsilon^2\mathbf{I}$. We remark that the observation error is about 1% and quite realistic compared to typical observations in e.g., atmospheric sciences.

For nowcasting, we also generate observations at a time $t_k - \delta_s$, where we choose $\delta_s = 0.02$, i.e., 1/6 of the time between two consecutive analysis steps. This would correspond to a 10-min nowcasting observation interval within an hourly data assimilation cycle, which is a realistic range compared for example to radar precipitation nowcasting in classical atmospheric forecasting systems. We test nowcasting for a lead time of $\gamma \in [0, 6]$, which corresponds to nowcasting up to one assimilation cycle or 1 h in the framework of typical atmospheric nowcasting.

In Fig. 2, the nowcasting is indicated by straight gray lines with black dots at the end, which indicate the nowcasted state at time $s = s_k + \gamma\delta_s$ used as data for the assimilation cycle. Blue dots show the ensemble mean of the first guess at analysis time, the red dots the mean of the analysis ensemble, observations are shown in black. The gray curves indicate the analysis ensemble; its variance is approximately the size of the observation error, i.e., 0.02.

We carry out our case study with $N_{\text{nat}} = 100$ steps of the analysis cycle. This corresponds to approximately 4–5 days of atmospheric convective scale data assimilation with an hourly assimilation cycle. We carry out ensemble data assimilation based on the LETKF as described in section 3. Spread control is based on the statistics of observations minus first guess (cf. Desroziers et al. (2005) and Li et al. (2009)) as also described in Potthast et al. (2019), given by

$$\mathbf{d}_{o-b} = \mathbf{y}^o - H(\mathbf{x}^b) = \mathbf{y}^o - H(\mathbf{x}^t) + H(\mathbf{x}^t) - H(\mathbf{x}^b) \approx \boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b, \tag{52}$$

with the true background state $\mathbf{x}^t$, the background state $\mathbf{x}^b$, the linearization $\mathbf{H}$ of H, the vector of observation errors $\boldsymbol{\epsilon}^o$, and the vector of background errors $\boldsymbol{\epsilon}^b$. If the observation errors and background errors are uncorrelated, we obtain

$$\mathbb{E}\left[\mathbf{d}_{o-b}\mathbf{d}^{\mathrm{T}}_{o-b}\right] = \mathbb{E}\left[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^{\mathrm{T}}\right] + \mathbf{H}\mathbb{E}\left[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^{\mathrm{T}}\right]\mathbf{H}^{\mathrm{T}}. \tag{53}$$

To estimate the inflation factor we substitute the expectation values of the background and observation error with the actual ensemble covariance matrix $\mathbf{P}^b$ multiplied by the inflation factor ρ and the nominal covariance of the observation error $\mathbf{R}$, respectively: $\mathbb{E}\left[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^{\mathrm{T}}\right] \approx \rho\mathbf{P}^b$ and $\mathbb{E}\left[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^{\mathrm{T}}\right] \approx \mathbf{R}$ resulting in

$$\mathbb{E}\left[\mathbf{d}_{o-b}\mathbf{d}^{\mathrm{T}}_{o-b}\right] \approx \mathbf{R} + \rho\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}. \tag{54}$$

By taking the *trace* $\mathrm{Tr}(\mathbf{A}) = \sum_{j=1}^m a_{jj}$ of the matrices on both sides, using $\mathrm{Tr}(\mathbf{A} + \tilde{\mathbf{A}}) = \mathrm{Tr}(\mathbf{A}) + \mathrm{Tr}(\tilde{\mathbf{A}})$, $\mathrm{Tr}\left(\rho\tilde{\mathbf{A}}\right) = \rho\mathrm{Tr}(\tilde{\mathbf{A}})$ and $\mathrm{Tr}(\mathbf{v}\mathbf{v}^{\mathrm{T}}) = \mathrm{Tr}(\mathbf{v}^{\mathrm{T}}\mathbf{v})$, the inflation factor ρ is estimated by

$$\rho = \frac{\mathbb{E}\left[\mathbf{d}^{\mathrm{T}}_{o-b}\mathbf{d}_{o-b}\right] - \mathrm{Tr}(\mathbf{R})}{\mathrm{Tr}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}})}. \tag{55}$$

The above factor is integrated into the code, estimated at each analysis step and smoothened over time with a decay factor of 0.8.

First, we study the system with different lead times for the nowcasted information. Figure 2 provides an impression of the dynamics of the nowcasting. We remark that $(s - s_k)/\delta_s = 6$ corresponds to the cycling interval of the data assimilation cycle. This means that the nowcasted state stays on a straight line, while the system dynamics follows a quarter of a circle of one of the wings of the butterfly. Clearly, there are strong differences between the linearly nowcasted state and the nonlinear dynamics of the Lorenz system. But this difference is also there for the nowcasting of simulated observations. So when $\mathbf{R}$ is chosen by (50), the assimilation of the nowcasted state integrates the derivatives of the system at the analysis time by (18), and it gives these derivatives some weight by the lead time of the nowcasted states.

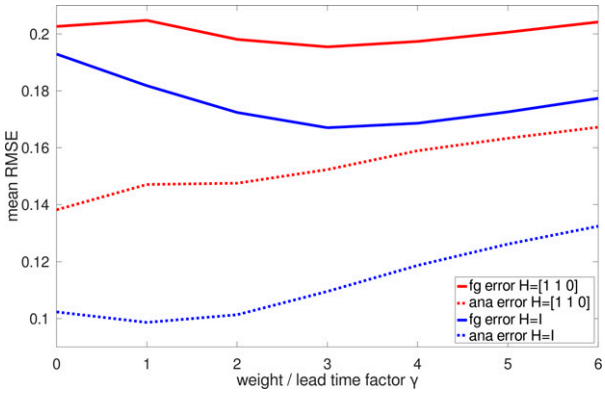

FIG. 5. For the Lorenz 63 model we display the mean first guess error (full red line) and the mean analysis error (dotted line) for the analysis cycle assimilating nowcasted information with different lead times γ. The blue lines display results for $H = I$, red lines display results for $H = (1, 1, 0)$. Here, we show results for weight or lead time factor $\gamma \in [0, 6]$, approximately for nowcasting lead times of zero to the analysis cycle interval.

Next, we evaluate the overall system performance in terms of the statistics of the first guess and analysis errors compared with the truth for the $N_{\text{nat}} = 100$ cycles for which we carry out the assimilation. Figure 5 displays the mean first guess and analysis errors for different lead times of the nowcasted information. Here, we take the mean error over all N_{nat} analysis time steps. The dotted curve shows the analysis error, the full line the first guess error, both compared to the nature run. We also carried out a standard data assimilation cycle with observations at t_k only, i.e., with the 3D version of the LETKF. The mean first guess error in this case was $e_b = 0.252\,88$ and $e_a = 0.121\,07$. When we take into account observations at t_k and $t_k - \delta_s$, we obtain $e_b = 0.229\,88$ and $e_a = 0.112\,53$. This corresponds to a 4D version of the LETKF, where information is employed at the correct point in time on the interval $[t_{k-1}, t_k]$ for the analysis at t_k. The improvement seems plausible, since we have more independent observations of the state in four-dimensional case. The four-dimensional case is also the starting point of nowcasted information, since in the case $s = s_k$ it coincides with using nowcasting evaluated at time $s_{k,1}$, the beginning of the graphs at $\gamma = 0$ in Fig. 5.

Figure 5 visualizes several important results, using blue lines for the observation operator $H = I$ and with red lines for $H = (1, 1, 0)$. First, the error value for $\gamma = 0$ is the case where we assimilate observations at t_k and t_k–δ_s independently. This corresponds to a 4D version with the active use of two time steps close to the end of the time interval under consideration for each analysis cycle. The use of nowcasted information reduces the forecast error significantly approximately for a nowcasting lead time up to 1–2 analysis cycles. For atmospheric data assimilation with hourly assimilation cycle this corresponds to nowcasting of about two hours. In dependence on γ, starting from $\gamma = 1$ usually the average first-guess forecast error first decreases with growing γ. After reaching a minimum it starts to grow with growing lead time of the nowcasted information.

Further, with the blue lines in Fig. 5, where $H = I$, we observe that the analysis error is minimal for $\gamma = 1$, which corresponds to the case where only observations at analysis time are assimilated. With growing γ the analysis error starts to grow, though still the forecast error decays until $\gamma = 3$. We note that we measure the analysis error in the L^2 norm at analysis time, while the assimilation functional now employs a norm which consists of the L^2 distance plus an L^2 distance on the nowcasted states. This norm resembles a weighted H^1 metric, which includes the derivative. This means that a good fit of the derivative might lead to a worse fit of the state itself, but to include the direction of motion leads to better forecasts. Displayed by the red lines in Fig. 5 we observe and nowcast the sum of the variables x_1 and x_2. Here, the assimilation error is best when the information at s_k and t_k is used. Again there is benefit for the forecast score when nowcasted observations are assimilated with $\mathbf{R}$ given by (50), as shown by the minimum of the bold curve at $\gamma = 3$.

6. Conclusions and further steps

This work studies the assimilation of nowcasted information into a classical data assimilation cycle. We employ the Localized Ensemble Transform Kalman Filter (LETKF) as assimilation method. Nowcasting is modeled as a linear extrapolation of observations based on two temporally close measurements. We compare the assimilation of observations at analysis time t_k with the assimilation of two observations at t_k and $s_k = t_k - \delta_s$ and the assimilation of observations at t_k plus nowcasted information with a nowcast lead time factor $\gamma = (s - s_k)/\delta_s$. We describe the integration of nowcasted information into the LETKF. The theory for the assimilation of nowcasted information has been worked out by calculating the transformed observation error covariance matrix under the assumption that observations at different time steps have an independent observation error with an observation error covariance matrix $\mathbf{R}_0$. When the same transformation is used for nowcasting of observations and the observation error covariance matrix $\mathbf{R}$, it has been shown theoretically that no forecasts improvements can be obtained. However, with a different matrix R we can observe improved forecast scores. A numerical study based both on an oscillator model and the popular Lorenz 63 model as a chaotic dynamical system shows the feasibility of the assimilation of nowcasted information with an ensemble data assimilation system as well as the equivalence or benefit of nowcasted information over the classical three-dimensional or four-dimensional assimilation of this information depending on the choices of the covariance matrix $\mathbf{R}$. We expect these results to lead to further research on the assimilation of nowcasted information in typical atmospheric forecasting systems on the convective scale. The results are very promising since they improve the forecast errors within the assimilation cycle significantly.

One key next step is the implementation and experimental testing of the techniques in a full data assimilation system for numerical weather prediction. For the Kilometer-scale Ensemble Data Assimilation System (KENDA) of the COSMO Consortium the calculation of observation equivalents, i.e., the

setup of the vectors **y** of Eq. (18) and the matrices **Y** defined in (31) is carried out during model runs and is passed to the data assimilation system by so-called *feedback files* in netCDF format. These form a natural interface to feed nowcasted information (31) into KENDA, when the nowcasted information together with their nowcasted simulated counterparts are stored in the feedback file format and read in the standard way. The 3D-volume radar observations are assimilated operationally in the regional model of DWD and are a good candidate for testing simplified nowcasting and its assimilation. Technically, the assimilation of more sophisticated nowcasting objects can be carried out in the same way and is in preparation.

Acknowledgments. The second author was funded by the RealPEP Research Group of Deutsche Forschungsgemeinschaft DFG, Grant 404442345, GZ PO 543/9-1, and SI 606/32-1.

REFERENCES

Agarwal, R. P., and D. O'Regan, 2008: *An Introduction to Ordinary Differential Equations.* Springer, 322 pp., https://doi.org/10.1007/978-0-387-71276-5.

Ambadan, J. T., and Y. Tang, 2009: Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems. *J. Atmos. Sci.*, **66**, 261–285, https://doi.org/10.1175/2008JAS2681.1.

Anandan, P., 1989: A computational framework and an algorithm for the measurement of visual motion. *Int. J. Comput. Vis.*, **2**, 283–310, https://doi.org/10.1007/BF00158167.

Anderson, B. D. O., and J. B. Moore, 2012: *Optimal Filtering.* Dover Books on Electrical Engineering Series, Dover, 357 pp.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 607–633, https://doi.org/10.1002/qj.2982.

Bechini, R., and V. Chandrasekar, 2017: An enhanced optical flow technique for radar nowcasting of precipitation and winds. *J. Atmos. Oceanic Technol.*, **34**, 2637–2658, https://doi.org/10.1175/JTECH-D-17-0110.1.

Bellon, A., and G. L. Austin, 1978: The evaluation of two years of real-time operation of a short-term precipitation forecasting procedure (SHARP). *J. Appl. Meteor.*, **17**, 1778–1787, https://doi.org/10.1175/1520-0450(1978)017<1778:TEOTYO>2.0.CO;2.

Bowler, E. H., N. C. E. Pierce, and A. Seed, 2004: Development of a precipitation nowcasting algorithm based upon optical flow techniques. *J. Hydrol.*, **288**, 74–91, https://doi.org/10.1016/j.jhydrol.2003.11.011.

Carrassi, A., and S. Vannitsem, 2010: Accounting for model error in variational data assimilation: A deterministic formulation. *Mon. Wea. Rev.*, **138**, 3369–3386, https://doi.org/10.1175/2010MWR3192.1.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: 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.

Engl, H. W., M. Hankle, and A. Neubauer, 2000: *Regularization of Inverse Problems.* Mathematics and Its Applications, Springer, 322 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342–1354, https://doi.org/10.1175/1520-0493(1997)125<1342:ADAFSN>2.0.CO;2.

——, 2009: *Data Assimilation: The Ensemble Kalman Filter.* Springer, 307 pp.

——, and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Foresti, L., I. V. Sideris, L. Panziera, D. Nerini, and U. Germann, 2018: A 10-year radar-based analysis of orographic precipitation growth and decay patterns over the Swiss alpine region. *Quart. J. Roy. Meteor. Soc.*, **144**, 2277–2301, https://doi.org/10.1002/qj.3364.

Germann, U., and I. Zawadzki, 2002: Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Mon. Wea. Rev.*, **130**, 2859–2873, https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2.

——, and ——, 2004: Scale-dependence of the predictability of precipitation from continental radar images. Part II: Probability forecasts. *J. Appl. Meteor.*, **43**, 74–89, https://doi.org/10.1175/1520-0450(2004)043<0074:SDOTPO>2.0.CO;2.

Glahn, B., A. D. Schnapp, J. E. Ghirardelli, and J.-S. Im, 2017: A LAMP–HRRR MELD for improved aviation guidance. *Wea. Forecasting*, **32**, 391–405, https://doi.org/10.1175/WAF-D-16-0127.1.

Golding, B. W., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteor. Appl.*, **5**, 1–16, https://doi.org/10.1017/S1350482798000577.

Goodliff, M., J. Amezcua, and P. J. V. Leeuwen, 2015: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity. *Tellus*, **67A**, 26928, https://doi.org/10.3402/tellusa.v67.26928.

Haiden, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber, 2011: The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the eastern alpine region. *Wea. Forecasting*, **26**, 166–183, https://doi.org/10.1175/2010WAF2222451.1.

Hodyss, D., 2011: Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation. *Mon. Wea. Rev.*, **139**, 3571–3588, https://doi.org/10.1175/2011MWR3558.1.

Hohti, H., J. Koistinen, P. Nurmi, E. Saltikoff, and K. Holmlund, 2000: Precipitation nowcasting using radar-derived atmospheric motion vectors. *Phys. Chem. Earth, Part B Hydrol. Oceans Atmos.*, **25**, 1323–1327, https://doi.org/10.1016/S1464-1909(00)00202-1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

——, and ——, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, https://doi.org/10.1256/qj.05.135.

——, and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

——, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, https://doi.org/10.1175/MWR-2864.1.

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.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Lei, J., and P. Bickel, 2011: A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Wea. Rev.*, **139**, 3964–3973, https://doi.org/10.1175/2011MWR3553.1.

Lei, L., D. R. Stauffer, S. E. Haupt, and G. S. Young, 2012: A hybrid nudging-ensemble Kalman filter approach to data assimilation. Part I: Application in the Lorenz system. *Tellus*, **64A** 18484, https://doi.org/10.3402/tellusa.v64i0.18484.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, https://doi.org/10.1002/qj.371.

Li, L., W. Schmid, and J. Joss, 1995: Nowcasting of motion and growth of precipitation with radar over a complex orography. *J. Appl. Meteor.*, **34**, 1286–1300, https://doi.org/10.1175/1520-0450(1995)034<1286:NOMAGO>2.0.CO;2.

Lorenc, A. C., and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, https://doi.org/10.1002/qj.49712657002.

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

Marzban, C., 2013: Variance-based sensitivity analysis: An illustration on the Lorenz'63 model. *Mon. Wea. Rev.*, **141**, 4069–4079, https://doi.org/10.1175/MWR-D-13-00032.1.

Nakamura, G., and R. Potthast, 2015: *Inverse Modeling: An Introduction to the Theory and Methods of Inverse Problems and Data Assimilation.* IOP Publishing, 509 pp., https://doi.org/10.1088/978-0-7503-1218-9.

Otkin, J. A., R. Potthast, and A. S. Lawless, 2020: Nonlinear conditional bias estimation for data assimilation. *SIAM J. Appl. Dyn. Syst.*, **20**, 299–332, https://doi.org/10.1137/19M1294848.

Potthast, R., and C. Welzbacher, 2018: Ultra rapid data assimilation based on ensemble filters. *Front. Appl. Math. Stat.*, **4**, 45, https://doi.org/10.3389/fams.2018.00045.

——, A. Walter, and A. Rhodin, 2019: A localized adaptive particle filter within an operational NWP framework. *Mon. Wea. Rev.*, **147**, 345–362, https://doi.org/10.1175/MWR-D-18-0028.1.

Pu, Z., and J. Hacker, 2009: Ensemble-based Kalman filters in strongly nonlinear dynamics. *Adv. Atmos. Sci.*, **26**, 373–380, https://doi.org/10.1007/s00376-009-0373-9.

Reich, S., and C. Cotter, 2015: *Probabilistic Forecasting and Bayesian Data Assimilation.* Cambridge University Press, 297 pp., https://doi.org/10.1017/CBO9781107706804.

Rinehart, R. E., 1981: A pattern recognition technique for use with conventional weather radar to determine internal storm motions. *Atmos. Technol.*, **13**, 119–134.

Ryu, S., G. Lyua, Y. Doc, and G. Lee, 2020: Improved rainfall nowcasting using Burgers equation. *J. Hydrol.*, **581**, 124140, https://doi.org/10.1016/j.jhydrol.2019.124140.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative ENKF for strongly nonlinear systems. *Mon. Wea. Rev.*, **140**, 1988–2004, https://doi.org/10.1175/MWR-D-11-00176.1.

Saltzman, B., 1962: Finite amplitude free convection as an initial value problem-I. *J. Atmos. Sci.*, **19**, 329–341, https://doi.org/10.1175/1520-0469(1962)019<0329:FAFCAA>2.0.CO;2.

Sideris, I. V., L. Foresti, D. Nerini, and U. Germann, 2020: Nowprecip: Localized precipitation nowcasting in the complex terrain of Switzerland. *Quart. J. Roy. Meteor. Soc.*, **146**, 1768–1800, https://doi.org/10.1002/qj.3766.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, https://doi.org/10.1175//2555.1.

Turner, B. J., I. Zawadzki, and U. Germann, 2004: Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE). *J. Appl. Meteor.*, **43**, 231–248, https://doi.org/10.1175/1520-0450(2004)043<0231:POPFCR>2.0.CO;2.

van Leeuwen, P. J., Y. Cheng, and S. Reich, 2015: *Nonlinear Data Assimilation.* Springer, 118 pp., https://doi.org/10.1007/978-3-319-18347-3.

Verlaan, M., and A. W. Heemink, 2001: Nonlinearity in data assimilation applications: A practical method for analysis. *Mon. Wea. Rev.*, **129**, 1578–1589, https://doi.org/10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2.

Vukicevic, T., and D. Posselt, 2008: Analysis of the impact of model nonlinearities in inverse problem solving. *J. Atmos. Sci.*, **65**, 2803–2823, https://doi.org/10.1175/2008JAS2534.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Wilson, J. W., and I. Edwin Kessler, 1963: Use of radar summary maps for weather analysis and forecasting. *J. Appl. Meteor.*, **2**, 1–11, https://doi.org/10.1175/1520-0450(1963)002<0001:UORSMF>2.0.CO;2.

Yang, S.-C., E. Kalnay, and B. Hunt, 2012: Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Mon. Wea. Rev.*, **140**, 2628–2646, https://doi.org/10.1175/MWR-D-11-00313.1.

Zhang, S., Z. Liu, A. Rosati, and T. Delworth, 2012: A study of enhancive parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus*, **64A**, 10963, https://doi.org/10.3402/tellusa.v64i0.10963.