Received: 5 February 2021 | Revised: 26 July 2021 | Accepted: 5 August 2021 | Published on: 30 September 2021

DOI: 10.1002/qj.4157

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Targeted covariance inflation for 3D-volume radar reflectivity assimilation with the LETKF

Klaus Vobig[1] | **Klaus Stephan**[1] | **Ulrich Blahak**[1] | **Kobra Khosravian**[1] | **Roland Potthast**[1,2]

[1]Deutscher Wetterdienst, Data Assimilation Unit, Offenbach, Germany

[2]Department of Mathematics, University of Reading, Berkshire, UK

Correspondence
K. Vobig, Deutscher Wetterdienst, Data Assimilation Unit, Frankfurter Street 135, 63067 Offenbach, Germany
Email: klaus.vobig@dwd.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

The local ensemble transform Kalman filter (LETKF) suggested by Hunt *et al.*, 2007 is a very popular method for ensemble data assimilation. It is the operational method for convective-scale data assimilation at Deutscher Wetterdienst (DWD). At DWD, based on the LETKF, three-dimensional volume radar observations are assimilated operationally for the operational ICON-D2. However, one major challenge for the LETKF is the situation where observations show precipitation (reflectivity) whereas all ensemble members do not show such reflectivity at a given point in space. In this case, there is no sensitivity of the LETKF with respect to the observations, and the analysis increment based on the observed reflectivity is zero. The goal of this work is to develop a targeted covariance inflation (TCI) for the assimilation of 3D-volume radar data based on the LETKF, adding artificial sensitivity and making the LETKF react properly to the radar observations. The basic idea of the TCI is to employ an additive covariance inflation as entrance point for the LETKF. Here, we construct perturbations to the simulated observation which are used by the core LETKF assimilation step. The perturbations are constructed such that they exhibit a correlation between humidity and reflectivity. This leads to a change in humidity in such a way that precipitation is more likely to occur. We describe and demonstrate the theoretical basis of the method. We then present a case study where targeted covariance inflation leads to a clear improvement of the LETKF and precipitation forecast. All examples are based on the German radar network and the ICON-D2 model over Central Europe.

KEYWORDS

covariance inflation, data assimilation, local ensemble transform Kalman filter, radar reflectivity

1 | INTRODUCTION

The initialization of dynamical models for forecasting is carried out using data assimilation techniques (e.g., Lorenc *et al.*, 2000; Kalnay, 2003; Evensen, 2009; Kleist *et al.*, 2009; Anderson and Moore, 2012; Nakamura and Potthast, 2015; Reich and Cotter, 2015; van Leeuwen *et al.*, 2015; Houtekamer and Zhang, 2016; Bannister, 2017). Modern ensemble data assimilation (Evensen, 1994; 2009; Houtekamer and Mitchell, 1998; 2001; 2005; Evensen

and van Leeuwen, 2000; Anderson, 2001; Whitaker and Hamill, 2002; Snyder and Zhang, 2003; Houtekamer *et al.*, 2005) employs an ensemble of states to dynamically estimate the covariances both between atmospheric variables of the model and between atmospheric variables and observations. A very popular method for ensemble data assimilation is the ensemble Kalman filter (Evensen, 2009), which has evolved into many versions and realizations, one of which is the local ensemble transform Kalman filter (LETKF) of (Hunt *et al.*, 2006).

In this work, we employ the kilometer-scale ensemble data assimilation (KENDA) data assimilation framework (Schraff *et al.*, 2016), an implementation of the LETKF following (Hunt *et al.*, 2006), in combination with the (icosahedral nonhydrostatic (ICON) model (Zängl *et al.*, 2015; Prill *et al.*, 2020) in the limited area mode ICON-D2 covering Central Europe with 2 km horizontal resolution. The 3D-Volume radar observations are obtained from the C-band radar network of the German Weather Service, complemented by the European radar composite. Model equivalents are given through an application of the Efficient Modular Volume scanning Radar Operator (EMVORADO) forward operator (Zeng *et al.*, 2016) and are operationally assimilated by KENDA. The operational KENDA system also includes latent heat nudging (LHN) based on the radar composite of the precipitation scan of the radar stations and used in addition to 3D-volume radar data assimilation.

A strong challenge to the assimilation of radar reflectivities within an ensemble data assimilation system is the occasional presence of large discrepancies between observed and simulated reflectivities in combination with a vanishing background reflectivity spread, that is, where the reflectivity of all ensemble members is vanishing. This overconfidence in the background system state then leads to a vanishing contribution of this specific observation to the LETKF increment, thus the LETKF fails to synchronize the model state with the "true" system state, that is, nature, with respect to this particular observation.

To make the LETKF more sensitive to observations in such cases, we implement a (formally) additive covariance approach that employs correlations with model variables related to the specific humidity. Based on these correlations, a certain contribution is added to the simulated reflectivity for each ensemble member and, therefore, the spread is increased. As the added piece is not static but highly dynamic and relies on correlations with model variables, we refer to our approach as *targeted covariance inflation* (TCI). Note that, while these inflated simulated reflectivities are, strictly speaking, artificial, they are not random as there are physical processes involved that manifest themselves in correlations on which the TCI is then based on. Regarding the issue of zero spread of reflectivity, Yokota et al. (Yokota *et al.*, 2018) and Dowell and Wicker (Dowell and Wicker, 2009) have also suggested approaches to resolve this issue. The approach suggested by Yokota et al. and our TCI approach are based on the same basic idea of adding reflectivity correlated to model variables. However, there are some important differences. Firstly, we use a quite different algorithm to calculate the inflated reflectivities for each ensemble member, in our case based on *empirically* derived correlations between reflectivities and model quantities. The algorithm for inflating the spread used by Yokota et al. (Yokota *et al.*, 2018) is based on a total differential of a parametrization of the reflectivity, where the parametrization depends on model variables. Secondly, our case study is performed for a different meteorological situation (a tornadic supercell in their case versus a convective event in ours). Thirdly, we employ a different numerical model, the ICON model of DWD. This is widely used within the Consortium for Small-scale Modeling (COSMO), and to study the effect in this framework is of interest by itself. Lastly, we use a different setup, viz. an hourly assimilation cycle. Meanwhile, the work of Dowell and Wicker (Dowell and Wicker, 2009) is based on an additive noise inflation method that adds *random* perturbations to the ensemble members for maintaining spread. Furthermore, we formulate the TCI concept to be applied only if certain conditions are fulfilled (zero spread and discrepancy between observations and simulations).

Section 2 describes the parts of our systems and algorithms under consideration, in particular the assimilation framework KENDA, the ICON model, and the radar forward operator EMVORADO. Section 3 covers the theory and implementation of the TCI approach. In Section 4, the TCI is applied within the context of a single-observation experiment, and using a case study we investigate its impact on the assimilation and how convective structures inserted by the data assimilation step are maintained by a subsequent model run. Section 5 provides a short summary and an outlook.

2 | THE KENDA SYSTEM, ICON, AND THE EMVORADO RADAR OPERATOR

Here, we describe the system components used for implementing and studying the TCI. Section 2.1 introduces the KENDA data assimilation system. Section 2.1.1 introduces the notation and algorithmic steps of the LETKF used for the formulation and implementation of the TCI. The convective-scale ICON system is described in Section 2.2. For assimilating volume radar observations, we employ the EMVORADO operator described in Section 2.3.

2.1 | Kilometer-scale ensemble data assimilation with KENDA

The framework of our investigation of targeted covariance inflation is the KENDA system (Schraff *et al.*, 2016) of the COSMO consortium.[1] It is an implementation of the ensemble Kalman filter on the convective scale and is used operationally at Deutscher Wetterdienst (DWD) as well as further members and users of the COSMO Consortium. Today, KENDA is operationally employed for the ICON-D2 model at DWD.

The KENDA system consists of several parts, which are important for the context of our investigation. It realizes a LETKF following (Hunt *et al.*, 2006), where a deterministic analysis is based on the Kalman gain for the analysis ensemble mean.

KENDA includes several tools for adaptive localization, multiplicative covariance inflation, relaxation to prior perturbations (RTPP) and prior spread (RTPS), adaptive observation error, and latent heat nudging (LHN) using radar-derived precipitation rates. For more details on KENDA and its components, we refer to (Schraff *et al.*, 2016).

2.1.1 | The LETKF

Here, we employ the LETKF as described by (Hunt *et al.*, 2007). Each analysis cycle starts with an ensemble of L members

$$\left\{\mathbf{x}_{k-1}^{a,\ell} \;:\; \ell = 1, 2, \dots, L\right\} \tag{1}$$

at time t_{k-1} with time index k, where each member represents an n-dimensional model state vector. The ensemble members are then propagated by the forecast model, and an ensemble of background model state vectors at time t_k is obtained $\{\mathbf{x}_k^{b,\ell} \;:\; \ell = 1, 2, \dots, L\}$.

The mean and covariance of this background ensemble at time t_k are given through the sample mean and covariance of the background ensemble

$$\overline{\mathbf{x}}^b = L^{-1}\sum_{\ell=1}^{L}\mathbf{x}^{b,\ell}, \tag{2}$$

$$\mathbf{P}^b = (L-1)^{-1}\mathbf{X}^b(\mathbf{X}^b)^T, \tag{3}$$

where $\mathbf{X}^b$ is an $n \times L$ matrix whose ℓth column contains the background ensemble perturbation $\mathbf{x}^{b,\ell} - \overline{\mathbf{x}}^b$. The analysis at time t_k constructs an ensemble $\left\{\mathbf{x}_k^{a,\ell} \;:\; \ell = 1, 2, \dots, L\right\}$ such that

$$\overline{\mathbf{x}}^a = L^{-1}\sum_{\ell=1}^{L}\mathbf{x}^{a,\ell}, \tag{4}$$

$$\mathbf{P}^a = (L-1)^{-1}\mathbf{X}^a(\mathbf{X}^a)^T, \tag{5}$$

where similarly to $\mathbf{X}^b$, $\mathbf{X}^a$ is an $n \times L$ matrix whose ℓth column contains $\mathbf{x}^{a,\ell} - \overline{\mathbf{x}}^a$. The increment of the LETKF is calculated within the column space of $\mathbf{X}^b$. A real-valued vector $\mathbf{w} \in \mathbb{R}^L$ corresponds to a model state $\mathbf{x} = \overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}$. The cost function of the LETKF is defined as

$$\tilde{J}(\mathbf{w}) = (L-1)\mathbf{w}^T\mathbf{w} + \left(\mathbf{y}^o - H(\overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - H(\overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})\right), \tag{6}$$

where $\mathbf{y}^o$ denotes the m-dimensional observation vector. The observation operator is linearized via the approximation

$$H(\overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}) \approx \overline{\mathbf{y}}^b + \mathbf{Y}^b\mathbf{w}. \tag{7}$$

Here, $\overline{\mathbf{y}}^b$ denotes the sample mean of the observation background ensemble $\{\mathbf{y}^{b,\ell} \;:\; \ell = 1, 2, \dots, L\}$, where $\mathbf{y}^{b,\ell} = \mathbf{H}(\mathbf{x}^{b,\ell})$ are the m-dimensional observation-space equivalents of $\mathbf{x}^{b,\ell}$, and $\mathbf{Y}^b$ denotes the corresponding $m \times L$ matrix whose ℓth column is made up of the perturbation $\mathbf{y}^{b,\ell} - \overline{\mathbf{y}}$. Using this linearization, we obtain the cost function

$$\tilde{J}^*(\mathbf{w}) = (L-1)\mathbf{w}^T\mathbf{w} + \left(\mathbf{y}^o - \overline{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}\right)^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \overline{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}\right). \tag{8}$$

As the cost function is formulated within the low-dimensional ensemble space, its minimum can be computed explicitly, which yields

$$\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b), \tag{9}$$

$$\tilde{\mathbf{P}}^a = \left((L-1)\mathbf{1} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\mathbf{Y}^b\right)^{-1}, \tag{10}$$

and the associated mean and covariance in model space are

$$\begin{aligned}\overline{\mathbf{x}}^a &= \overline{\mathbf{x}}^b + \mathbf{X}^b\overline{\mathbf{w}}^a \\ &= \overline{\mathbf{x}}^b + \mathbf{X}^b\tilde{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b),\end{aligned} \tag{11}$$

$$\mathbf{P}^a = \mathbf{X}^b\tilde{\mathbf{P}}^a(\mathbf{X}^b)^T. \tag{12}$$

The Kalman gain $\mathbf{K} = \mathbf{X}^b\tilde{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}$ determines the increments for the analysis mean when supplied with the innovation $(\mathbf{y}^o - \overline{\mathbf{y}}^b)$. In our implementation of the LETKF,

[1] the Consortium for Small-Scale Modeling, including Italy, Switzerland, Romania, Greece, Russia, Poland, and Germany.

that is, within the context of KENDA, there is an additional *deterministic* run whose background and analysis state is denoted as $\mathbf{x}^b$ and $\mathbf{x}^a$, respectively. The analysis of this deterministic run is given through

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^b)), \tag{13}$$

such that the increment for the deterministic run is governed by the Kalman gain and the innovation with respect to the deterministic background. The ensemble members are constructed by

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{W}^a, \tag{14}$$

$$\mathbf{W}^a = \left((L-1)\tilde{\mathbf{P}}^a\right)^{(1/2)}, \tag{15}$$

where a symmetric square root is employed. The ensemble members are then given by

$$\mathbf{x}^{a,\ell} = \overline{\mathbf{x}}^b + \mathbf{X}^b(\overline{\mathbf{w}}^a + \mathbf{W}^a_\ell), \tag{16}$$

with the ℓth column $\mathbf{W}^a_\ell$ of $\mathbf{W}^a$, illustrating that each analysis ensemble member is a linear combination of the background ensemble members.

In the case where all simulated observations are identical (e.g., when no reflectivity is simulated by all ensemble members), the matrix $\mathbf{Y}$ is zero, $\mathbf{W}^a$ is the identity matrix, and $\mathbf{w}^a = 0$. In this case, the analysis ensemble is identical to the background ensemble and observations do not have an effect on the analysis.

An important ingredient for a successful application of ensemble Kalman filter approaches is localization (Hunt *et al.*, 2006), which confines the inclusion of observations to a local region around each analysis grid point only. KENDA implements localization by treating each analysis grid point separately and scaling the entries of the inverse observation–error covariance matrix $\mathbf{R}^{-1}$ with respect to the distance of the observations from the analysis grid point. The scaling is based on the Gaspari–Cohn function (which is similar to a Gaussian but falls to zero beyond a certain range) with a certain horizontal and vertical localization range.

It is well known (e.g., Anderson and Anderson, 1999; Ott *et al.*, 2004; Hunt *et al.*, 2006) that a pure implementation of the Kalman filter equations underestimates the uncertainties of its background state (or, equivalently, past observations). Several approaches have been suggested to artificially inflate the background covariance. *Multiplicative inflation* (Anderson and Anderson, 1999) multiplies the background covariance matrix by a specific factor larger than one, *additive inflation* (Ott *et al.*, 2004; Hunt *et al.*, 2006) usually adds a certain (but small) multiple of the identity matrix to the background or analysis covariance matrix, and *relaxation to prior perturbation* (Zhang *et al.*, 2004) inflates the analysis covariance by relaxing each analysis ensemble perturbation towards its corresponding background ensemble perturbation. In KENDA, all these methods are available and are used in the operational setup. However, they do not help to solve the problem of zero spread when all ensemble members are identical in simulating a particular observation.

As a final part of KENDA, we need to discuss latent heat nudging (LHN). The LHN component of KENDA allows the assimilation of radar-derived precipitation rates. The LHN approach is based on the assumption that the precipitation rate near the surface is approximately proportional to the release of latent heat. This latent heat is then accounted for via the introduction of temperature increments into the model. Additionally, there is an increment in specific humidity to ensure that the relative humidity is conserved. The increments obtained from the LHN approach are horizontally local with respect to the underlying precipitation rates and applied during the model run. For more information on the concepts and details of the LHN approach, see (Stephan *et al.*, 2008).

2.2 | The convective-scale atmospheric model ICON-LAM

The ICON modeling framework (Zängl *et al.*, 2015; Prill *et al.*, 2020) is the numerical weather prediction and climate modeling system developed by a consortium of institutions and weather services, in particular Deutscher Wetterdienst (DWD) and the Max-Planck-Institut for Meteorology (MPI-M). The ICON model runs operationally at the DWD on a global scale as well as on a European subdomain (ICON-EU). The convection-permitting model setup ICON-D2, whose model domain coincides roughly with that of COSMO-D2, is currently becoming the operational model of DWD, replacing the COSMO model (Baldauf *et al.*, 2011).

The system of equations solved by the ICON model is based on a specific set of prognostic variables, and broadly speaking, a two-component system is assumed, consisting of dry air and water that may occur in all three phases including drops and ice particles. More specifically, the system of equations contains the horizontal velocity component normal to triangle edges v_n, the vertical wind component w, the virtual potential temperature θ_v, the total density of air mixture $\rho = \sum_k \rho_k$, and mass fractions $q_k = \rho_k/\rho$. Note that the index k occurring within the density ρ_k and mass fraction q_k refers to dry air ($k = d$), water vapor ($k = v$), cloud water ($k = c$), cloud ice ($k = i$), rain ($k = r$), snow ($k = s$), and graupel ($k = g$).

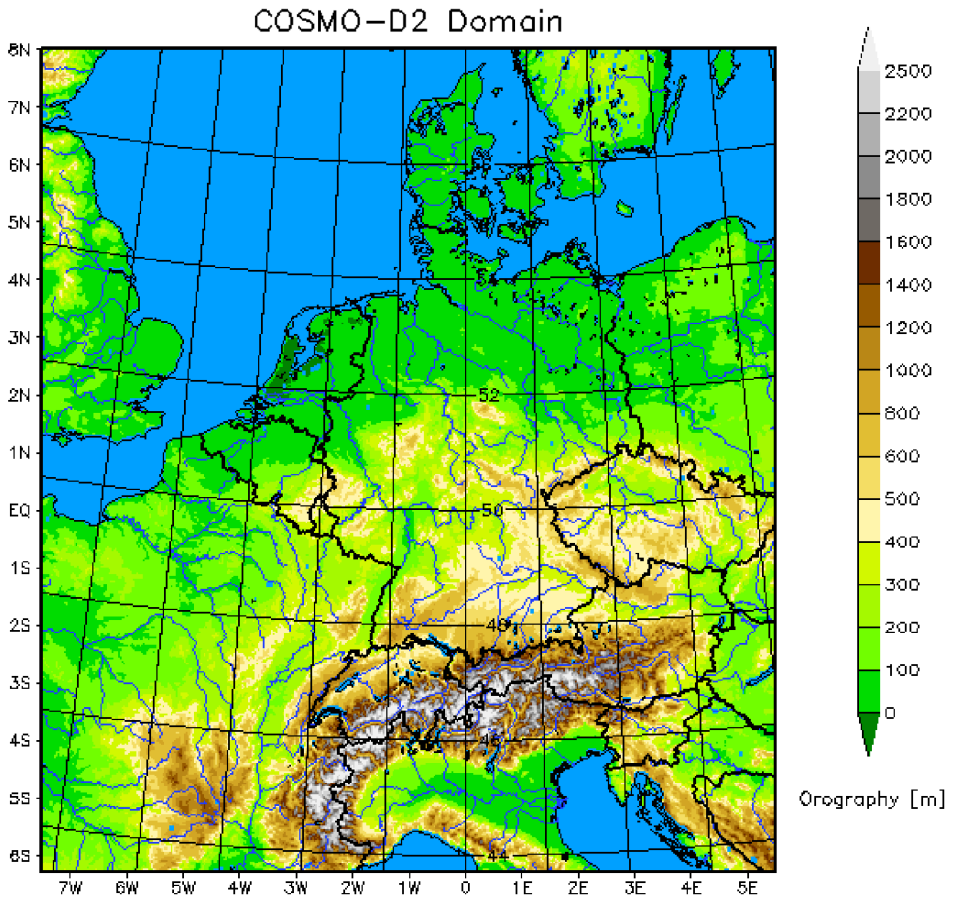

FIGURE 1 Model domain of the COSMO-D2-KENDA system. The ICON-D2-KENDA system has approximately the same extent. The map is given in rotated coordinates [Colour figure can be viewed at wileyonlinelibrary.com]

Horizontally, the ICON model employs an unstructured triangular grid, and vertically, it defines a specific number of levels $\{l_i\}$, of which the lowest levels, that is, the ones closest to the ground, are terrain following whereas all higher levels gradually shift to levels of constant height. The levels are indexed from top to bottom such that l_1 is the highest level and l_{i+1} is below l_i.

We conduct ICON simulations in the limited area mode configuration ICON-D2, with a domain displayed in Figure 1. The ICON-D2 configuration uses a model resolution of approximately 2 km and has 65 vertical levels, and lateral boundary conditions are provided by ICON-EU simulations. For an in-depth discussion of the ICON model, see (Prill *et al.*, 2020).

2.3 | 3D-volume radar reflectivity simulation by EMVORADO

The observations used for the convective-scale configuration ICON-D2 of ICON include 3D-volume radar observations (Bick *et al.*, 2016). In the present work, two different sources for radar observations are used. On the one hand, we include the observations as measured by the German Radar Network by 3D volume radar data. The German Radar Network consists of 17 dual-polarization C-band Doppler radar stations completely covering Germany, as shown in Figure 2. The scanning strategy for 3D volume scans of each radar station, which is also depicted in Figure 3, consists of an azimuthal 360° sweep with a resolution of 1° at a total of ten elevation angles ranging from 0.5° to 25° and a radial distance of up to 180 km with a radial resolution of 1 km. Further, data supplied via the OPERA network, including the precipitation scan of the German Radar Network, enter the LHN mechanism.

To assimilate radar observations, synthetic radar observations are derived from these model variables based on the forward operator EMVORADO (Zeng *et al.*, 2016), yielding Doppler velocity and reflectivity in its current single-polarization implementation. EMVORADO yields simulated radar observations in observation space, that is, for each radar observation, a corresponding model equivalent is estimated. Several sophisticated physical aspects associated with radar measurements are taken into account by EMVORADO, in particular beam bending/broadening/shielding, Doppler velocity with fall speed and reflectivity weighting, attenuated reflectivity, and detectable signal. For more information and details on EMVORADO, see (Zeng *et al.*, 2016).

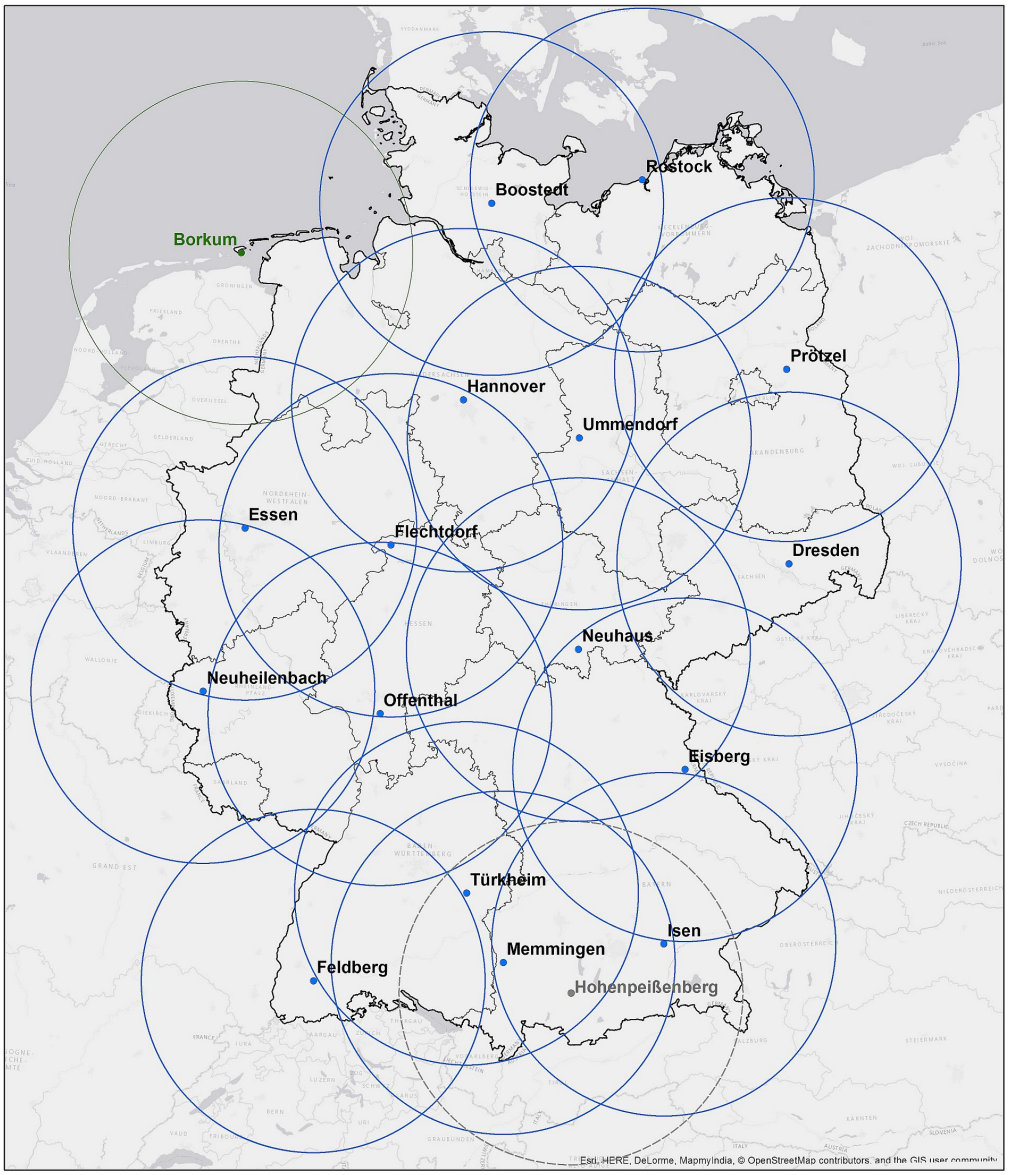

FIGURE 2 The German Radar Network, where blue dots represent the sites of the operational radar stations, while grey dots indicate radar stations for quality assurance and research purposes. The blue circles indicate a range of 150 km. The single green dot (Borkum) indicates an intermediate radar site without polarimetric observables [Colour figure can be viewed at wileyonlinelibrary.com]

3 | TCI FOR LETKF

In the following, we introduce and discuss a TCI approach for improving the assimilation of 3D radar data. The TCI adds a certain part to the background uncertainty covariance matrix and, therefore, is a type of additive covariance inflation (ACI). However, unlike a "traditional" additive covariance inflation, which selects random perturbations based on climatological covariances, the TCI employs particular targeted correlations with selected model variables for calculating the perturbations added to the background covariance matrix in observation space. This approach increases the uncertainty of the background in such a way that the additive increments use correlations between simulated reflectivities and model variables, such that these lead to a controlled change of the model variables in an ensemble data assimilation step.

3.1 | Theory and limitations of TCI

In the following, we denote the modeled (background) reflectivity at the spatial point $\mathbf{r} = (\lambda, \phi, h)$ of the ℓth ensemble member as $Z_{\mathbf{r}}^{(\ell)}$, where λ, ϕ, and h refer to longitude, latitude, and height above sea level. The corresponding ensemble of simulated reflectivities is given by $\mathbf{Z_r} = \{Z_{\mathbf{r}}^{(\ell)}\}$, the (background) reflectivity of the deterministic member by $Z_{\mathbf{r}}^{\mathrm{d}}$, and the reflectivity analysis and increment of the deterministic member by $Z_{\mathbf{r}}^{\mathrm{d,ana}}$ and $Z_{\mathbf{r}}^{\mathrm{d,inc}}$. The time information will usually be suppressed as it will be assumed that all quantities are given at assimilation time – if not stated otherwise.

The goal of the TCI is to overcome the problems associated with ensembles whose spread with respect to the reflectivity in a certain spatial region is zero or small, that is, $\sigma[\mathbf{Z_r}] \ll 1$. When Y^b is zero or small, the LETKF

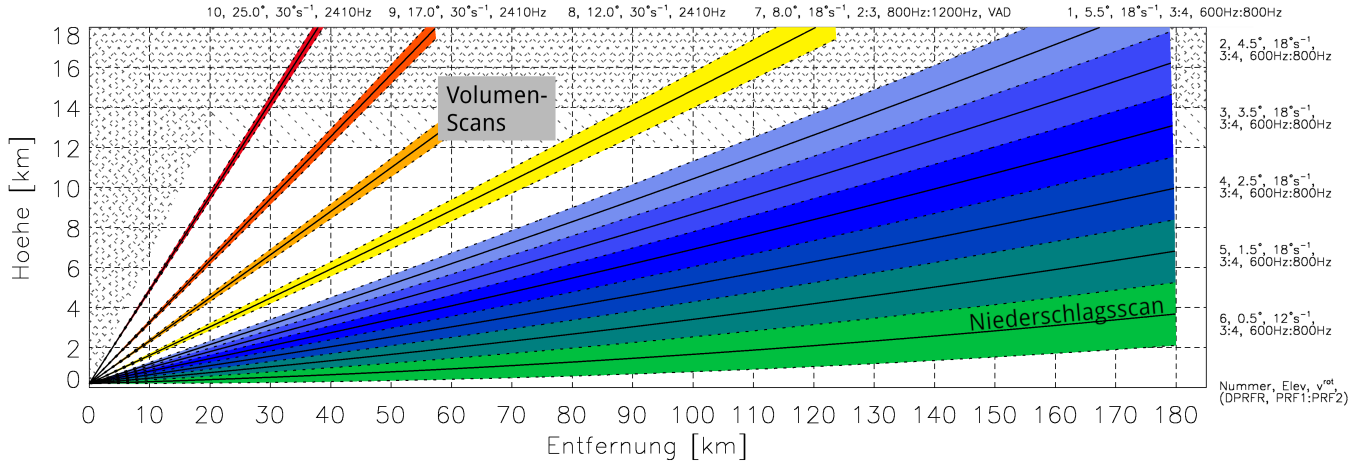

FIGURE 3 Scanning strategy of the German radar network, showing the range of each of the ten fixed elevations as well as of the terrain-following precipitation scan [Colour figure can be viewed at wileyonlinelibrary.com]

analysis (9)–(16) will show no or only very small analysis increments.

To overcome this deficiency, we inflate the background covariance in observation space based on correlations between Z and particular *model* variables denoted as Ψ at particular positions $\mathbf{r}$. Formally, the inflated reflectivities, denoted as Z' here, are then defined via

$$Z'^{(i)}_{\mathbf{r}} = Z^{(i)}_{\mathbf{r}} + \alpha_{\mathrm{TCI}} \left(\Psi^{(i)}_{\mathbf{r}} - \mu[\mathbf{\Psi}_{\mathbf{r}}] \right), \tag{17}$$

where α_{TCI} serves as a scaling factor for the strength of the TCI. As a consequence of this definition, the inflated reflectivity Z' has the following properties:

$$\mu[\mathbf{Z'}_{\mathbf{r}}] = \mu[\mathbf{Z}_{\mathbf{r}}], \tag{18}$$

$$\mathrm{var}[\mathbf{Z'}_{\mathbf{r}}] = \mathrm{var}[\mathbf{Z}_{\mathbf{r}}] + \alpha^2_{\mathrm{TCI}} \mathrm{var}[\mathbf{\Psi}_{\mathbf{r}}] + 2\alpha_{\mathrm{TCI}} \mathrm{cov}[\mathbf{Z}_{\mathbf{r}}, \mathbf{\Psi}_{\mathbf{r}}], \tag{19}$$

where $\mu[\cdot]$, $\mathrm{var}[\cdot]$, and $\mathrm{cov}[\cdot,\cdot]$ denote the mean, variance, and covariance, respectively.

In the following, we assume that the reflectivity is vanishing at our point of interest $\mathbf{r}_{so}$ for all ensemble members and the deterministic member, that is, $Z^{\mathrm{d}}_{\mathbf{r}_{so}} = Z^{(i)}_{\mathbf{r}_{so}} = 0$, which leads to

$$Z'^{(i)}_{\mathbf{r}_{so}} = \alpha_{\mathrm{TCI}} \left(\Psi^{(i)}_{\mathbf{r}_{so}} - \mu[\mathbf{\Psi}_{\mathbf{r}_{so}}] \right), \tag{20}$$

$$\mu[\mathbf{Z'}_{\mathbf{r}_{so}}] = 0, \tag{21}$$

$$\mathrm{var}[\mathbf{Z'}_{\mathbf{r}_{so}}] = \alpha^2_{\mathrm{TCI}} \mathrm{var}[\mathbf{\Psi}_{\mathbf{r}_{so}}]. \tag{22}$$

In this case, the variances of Z' and Ψ are directly proportional to each other and scaled by α^2_{TCI}.

3.2 | Single-Obs and increments

To study the effects of the TCI in more detail, we carry out a single-observation experiment; that is, a single observation at a specific spatial coordinate $\mathbf{r}_{so}$ is assimilated. We choose the reflectivity Z at a given point in space where the LETKF has its limitation. Assimilating one scalar value of Z, increments of other variables $\Omega_{\mathbf{r}}$ produced by the LETKF are solely based on correlations of the form $\rho[\mathbf{Z}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}]$.

For a single-observation experiment, the LETKF equation for obtaining the analysis state vector can be significantly simplified. We are studying a reformulation of our problem with a one-dimensional observation space whose elements are associated with the observed reflectivity $\mathbf{y}^o = Z^o_{\mathbf{r}_{so}}$ and a two-dimensional model space whose two dimensions are associated with the modeled reflectivity $\mathbf{Z}_{\mathbf{r}_{so}}$ at the spatial position of the single observation—serving as the model equivalent for the single observation—and another for the most part arbitrary model quantity $\Omega_{\mathbf{r}}$. The (deterministic) background state vector may then be written as $\mathbf{x}^b = \left(Z^{\mathrm{d}}_{\mathbf{r}_{so}} \; \Omega^{\mathrm{d}}_{\mathbf{r}} \right)^T$, and the observation operator connecting the model and observation space with each other takes the simple form $\mathbf{H} = (1\ 0)$. Employing Equation (13), we obtain

$$\begin{pmatrix} Z^{\mathrm{d,inc}}_{\mathbf{r}_{so}} \\ \Omega^{\mathrm{d,inc}}_{\mathbf{r}} \end{pmatrix} = \begin{pmatrix} \mathrm{var}[\mathbf{Z}_{\mathbf{r}_{so}}] \\ \mathrm{cov}[\mathbf{Z}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}] \end{pmatrix} (R + \mathrm{var}[\mathbf{Z}_{\mathbf{r}_{so}}])^{-1} (Z^o_{\mathbf{r}_{so}} - Z^{\mathrm{d}}_{\mathbf{r}_{so}}), \tag{23}$$

where we employed the Sherman–Morrison formula for analytically calculating the inverse appearing within the expression for the Kalman gain matrix. Using Equations (20) and (22) for expressing covariances involving the variable Z in terms of covariances involving Ψ, we obtain the following expression:

$$\Omega_{\mathbf{r}}^{\mathrm{d,inc}} = \frac{\alpha_{\mathrm{TCI}}\ \mathrm{cov}[\mathbf{\Psi}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}]}{R + \alpha_{\mathrm{TCI}}^2\ \mathrm{var}[\mathbf{\Psi}_{\mathbf{r}_{so}}]}(Z^o_{\mathbf{r}_{so}} - Z^{\mathrm{d}}_{\mathbf{r}_{so}}) \tag{24}$$

$$= \frac{\alpha_{\mathrm{TCI}}\ \sigma[\mathbf{\Psi}_{\mathbf{r}_{so}}]}{R + \alpha_{\mathrm{TCI}}^2\ \mathrm{var}[\mathbf{\Psi}_{\mathbf{r}_{so}}]}\rho[\mathbf{\Psi}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}]\sigma[\mathbf{\Omega}_{\mathbf{r}}](Z^o_{\mathbf{r}_{so}} - Z^{\mathrm{d}}_{\mathbf{r}_{so}}) \tag{25}$$

$$= \mathcal{F}(\alpha_{\mathrm{TCI}}, R, \sigma[\mathbf{\Psi}_{\mathbf{r}_{so}}])\rho[\mathbf{\Psi}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}]\sigma[\mathbf{\Omega}_{\mathbf{r}}](Z^o_{\mathbf{r}_{so}} - Z^{\mathrm{d}}_{\mathbf{r}_{so}}), \tag{26}$$

where we define

$$\mathcal{F}(\alpha_{\mathrm{TCI}}, R, \sigma[\mathbf{\Psi}_{\mathbf{r}_{so}}]) \equiv \frac{\alpha_{\mathrm{TCI}}\ \sigma[\mathbf{\Psi}_{\mathbf{r}_{so}}]}{R + \alpha_{\mathrm{TCI}}^2\ \mathrm{var}[\mathbf{\Psi}_{\mathbf{r}_{so}}]}, \tag{27}$$

and employ the definition of the correlation coefficient, that is,

$$\mathrm{cov}[\mathbf{\Psi}_{\mathbf{r}}, \mathbf{\Omega}_{\mathbf{r}}] = \rho[\mathbf{\Psi}_{\mathbf{r}}, \mathbf{\Omega}_{\mathbf{r}}]\sigma[\mathbf{\Psi}_{\mathbf{r}}]\sigma[\mathbf{\Omega}_{\mathbf{r}}].$$

While the specific values for the correlation coefficient $\rho[\mathbf{\Psi}_{\mathbf{r}_{so}}, \mathbf{\Omega}_{\mathbf{r}}]$, the variance $\mathrm{var}[\mathbf{\Psi}_{\mathbf{r}_{so}}]$, and the innovation $(Z^o_{\mathbf{r}_{so}} - Z^{\mathrm{d}}_{\mathbf{r}_{so}})$ are usually fixed input parameters, that is, determined through the model states and observations, the increment $\Omega_{\mathbf{r}}^{\mathrm{d,inc}}$ and, thereby, the impact of the TCI may still be tuned to specific needs via variation of the observation error R and the TCI scaling factor α_{TCI} as demonstrated by Figure 4.

Additionally, a localization scheme is applied by scaling the inverse of the observation covariance matrix $\mathbf{R}^{-1}$ with the Gaspari–Cohn function as already mentioned in Section 2.1.1. The Gaspari–Cohn function $f_h^{\mathrm{g.c.}}(d)$ gradually suppresses a contribution to an increment with respect to the distance d from the associated observation. The localization is carried out in vertical and horizontal direction independently, and the strength of the localization may be controlled via the localization range h where $f_h^{\mathrm{g.c.}}(d \geq 2h) = 0$ holds. In the context of a single-observation experiment, the effect of the localization can be written as

$$\overline{\Omega}^{\mathrm{inc}}(\mathbf{r}) \rightarrow \overline{\Omega}^{\mathrm{inc}}(\mathbf{r})\, f_{h_v}^{\mathrm{g.c.}}(|(\mathbf{r} - \mathbf{r}_{so})_v|)\ f_{h_h}^{\mathrm{g.c.}}(|(\mathbf{r} - \mathbf{r}_{so})_h|), \tag{28}$$

where $(\mathbf{r} - \mathbf{r}_{so})_v$ and $(\mathbf{r} - \mathbf{r}_{so})_h$ denote the vertical and horizontal component of $\mathbf{r} - \mathbf{r}_{so}$, respectively.

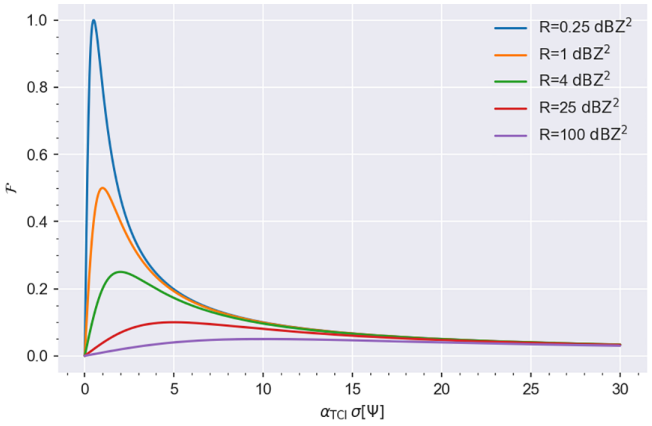

FIGURE 4 Using several values for the observation error R (shown in different colors) as a parameter, this plot displays the dependence of $\mathcal{F}$ defined in Equation (27) on $\alpha_{\mathrm{TCI}}\ \sigma[\mathbf{\Psi}_{\mathbf{r}_{so}}]$ [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Statistical evaluation for variable selection

To provide a sound basis of artificial targeted correlations, we carried out a series of statistical evaluations on correlations between reflectivity and model variables on the convective scale, whose details lie beyond the scope of this work. In summary, we found that reflectivity and humidity are only weakly correlated with each other, but there are stronger correlations between Z and the vertically integrated specific humidity, denoted as q_v^{int}:

$$q_v^{\mathrm{int}}(\lambda, \phi, l_0, l_1, \beta) \equiv \int_{\mathcal{A}} \mathrm{d}\lambda' \mathrm{d}\phi'\, f_\beta(\lambda' - \lambda, \phi' - \phi) \int_{h(l_0)}^{h(l_1)} q_v(\lambda', \phi', h)\mathrm{d}h. \tag{29}$$

The boundaries $h(l_0)$ and $h(l_1)$ of this integral correspond to the heights of the ICON layers l_0 and l_1, respectively, between which the numerical integration takes place. Furthermore, to account for time shifts of the correlations between Z and q_v^{int} leading to horizontal shifts of these two variables, a running average is applied to the vertically integrated fields. Implemented via a convolution using a rectangular function with respect to both arguments of f_β, the strength β of this running average controls the width of f_β, that is, the range within which f_β falls to zero.

In Figure 5, the correlation between $\delta Z^{(i)} \equiv Z^{(i)} - \mu[\mathbf{Z}]$ and $(\delta q_v^{\mathrm{int}})^{(i)} \equiv (q_v^{\mathrm{int}})^{(i)} - \mu[\mathbf{q}_{\mathbf{v}}^{\mathbf{int}}]$ is depicted by showing (q_v^{int}, Z) data pairs for all 40 ensemble members and all spatial positions of a specific area. The data pairs used for this correlation study lie within a domain with $8.5° < \lambda < 12.0°$ and $50° < \phi < 55°$ at 1200 UTC on June 3, 2019. We

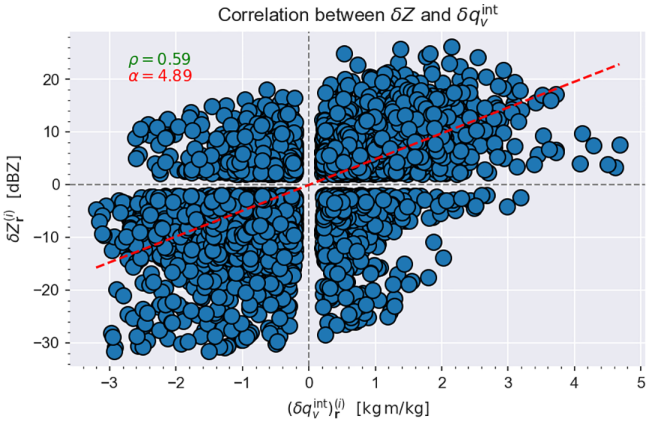

FIGURE 5 For all spatial points of a specific area bounded by $8.5° < \lambda < 12.0°$ and $50° < \phi < 55°$ and for all 40 ensemble members, the relationship between $\delta Z^{(i)} \equiv Z^{(i)} - \mu[\mathbf{Z}]$ and $(\delta q_v^{\text{int}})^{(i)} \equiv (q_v^{\text{int}})^{(i)} - \mu[\mathbf{q_v^{int}}]$ is shown. Note that we impose certain restrictions on the inclusion of data: Only radar data at an elevation angle of 0.5° are included here, and only data pairs fulfilling $\mu[\mathbf{Z_r}] > 10$ dBZ and $\sigma[\mathbf{Z_r}] > 8$ dBZ are included. Based on this data, the (nonlocal) correlation coefficient then amounts to 0.59. The red dashed line indicates the result of a least-squares fit using a linear function. For the slope of this linear function, we obtain $\alpha = 4.89$ dBZkgkg^{-1}m^{-1}. Note that we use $l_0 = 42$, $l_1 = 1$, and $\beta \approx 21$ km for the calculation of q_v^{int} here, where these specific values optimize the correlation coefficient [Colour figure can be viewed at wileyonlinelibrary.com]

chose this specific area and time as a basis of our statistical considerations as there are various convective events taking place, that is, the kind of meteorological situation whose dynamics we are trying to capture here.

For the calculation of q_v^{int}, we employ $l_0 = 42$, $l_1 = 1$, and $\beta \approx 21$ km, where these specific values are determined through a maximization of the correlation coefficient. Regarding the specific domain this statistical evaluation is based on, the average height of the model level l_0 amounts to roughly 2500 m, and the height of the level l_1, which also happens to represent the model top, amounts to 22,000 m. Note that these specific values for l_0, l_1, and β are now used as default values throughout the rest of this work. Figure 5 shows that the two variables Z and q_v^{int} are positively correlated. Performing a least-squares fit using a linear function, the slope of this linear function amounts to $\alpha = 4.89$ dBZkgkg^{-1}m^{-1} which serves as a guideline for the specific choice for the TCI scaling factor α_{TCI} later on.

Note that the statistical study is based on radar data at an elevation of only 0.5° whereas radar data at several elevations of up to 12° are assimilated in an operational setup. By including radar data only at this single elevation, we are trying not to overcomplicate things at first and we keep the integral for the vertically integrated q_v as simple as possible (as the heights we integrate over should mostly lie above the radar observation, we choose the lowest available elevation of 0.5°). There is, however, a slight impact of radar elevations: Alternatively using the elevation 1.5° as a basis for our statistical evaluation instead of 0.5°, the coefficient α amounts to 4.26 dBZ kg kg^{-1} ·m^{-1} and the correlation coefficient ρ amounts to 0.64. Even though there is this slight impact of the radar elevation on the result of the statistical evaluation of correlations, we consider the inclusion of data at an elevation of only 0.5° as sufficient for the purpose of this work.

Regarding future applications of the TCI, the aim to improve the description of deep convection could potentially require observations from higher radar elevations. We plan to study whether the possible issue of a small spread of the integrated q_v for greater values of l_0 as well as a decline of the correlation between reflectivities at higher elevations and the integrated q_v may then be resolved by using additional variables as a basis for the TCI. However, we want to leave these further, more complicated setups as a subject for future research.

4 | NUMERICAL RESULTS

Here, we evaluate the use of targeted covariance inflation on 3D-volume radar data assimilation for ICON-D2 by focusing on a case study. Section 4.1 describes the setup of the case study. The implementation of TCI is described in Section 4.2. The evaluation of LETKF increments is carried out in Section 4.3. Forecast impact is investigated in Section 4.4, and a comparison with LHN is carried out in Section 4.5.

4.1 | Case study setup

To study the effects of the TCI, we performed single-observation experiments assimilating a single radar reflectivity observation at $\mathbf{r}_{so} = (51.60°, 8.35°, 1{,}035\,m)$. This observation is obtained from the radar station Flechtdorf located at $(51.31°, 8.80°, 551\,m)$ and radar beam elevation angle of 0.5°. Thus, the radar station and single observation have a horizontal distance of about 45 km. The assimilation takes place on June 3, 2019 at $t_0 = 1200$ UTC.

Evaluating the reflectivity fields displayed in Figure 6, the choice of this specific location can be motivated. The radar composite for Z^o shows an isolated cell approximately centered around $\mathbf{r}_{so}$, which is indicated by red circles in the plots. However, the deterministic simulation lacks this particular cell completely, and as indicated by the vanishing ensemble spread of simulated reflectivities $\sigma[\mathbf{Z}]$, also none of the ensemble members contains simulated reflectivities in this region. As a result of the lack of ensemble spread, also the assimilation via the LETKF at

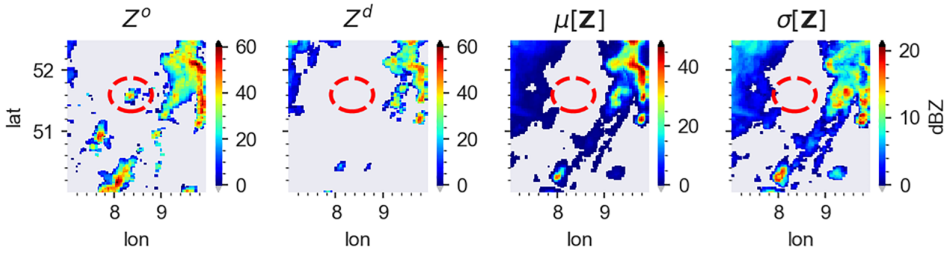

FIGURE 6 Depiction of radar composites at an elevation angle of 0.5° for Z^o, Z^d, $\mu[\mathbf{Z}]$, and $\sigma[\mathbf{Z}]$ on June 3, 2019 at $t_0 = 1200$ UTC. A specific region with an observed but not simulated cell is indicated via a red circle [Colour figure can be viewed at wileyonlinelibrary.com]

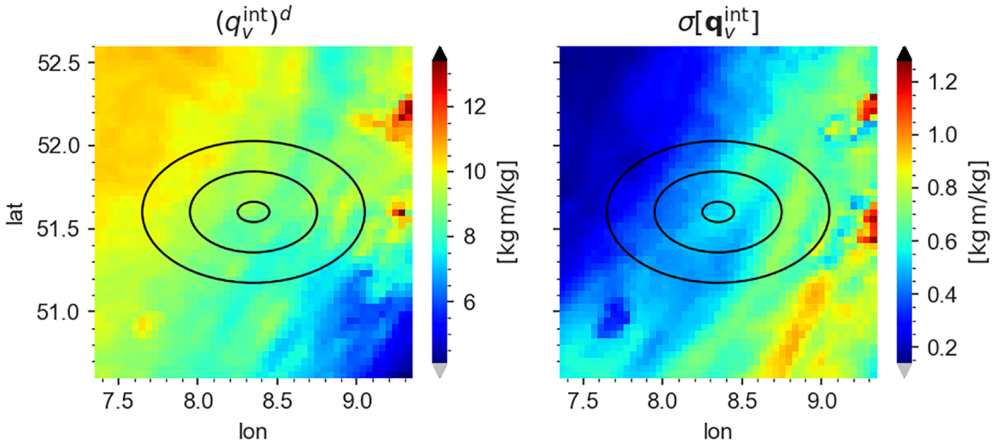

FIGURE 7 Depiction of $q_v^{\text{int}^d}$ and $\sigma[\mathbf{q}_\mathbf{v}^{\mathbf{int}}]$ at t_0. The fields for q_v^{int} are calculated by vertically integrating and horizontally averaging the model fields for q_v as discussed in Section 3.3. The position $\mathbf{r}_{so}$ of the single observation is indicated by concentric circles [Colour figure can be viewed at wileyonlinelibrary.com]

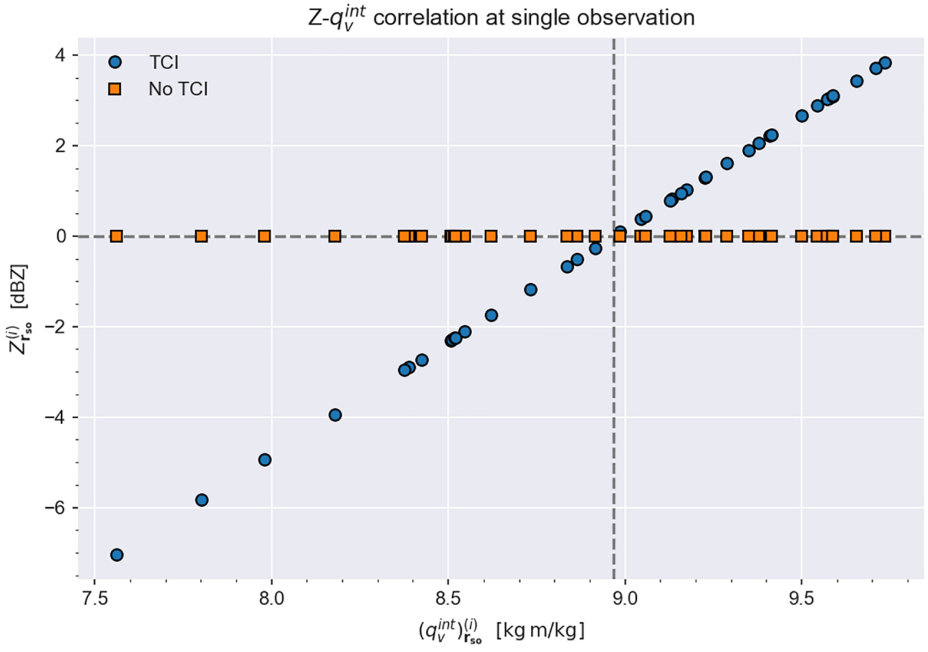

FIGURE 8 Depiction of $((q_v^{\text{int}})^{(i)}_{\mathbf{r}_{so}}, Z^{(i)}_{\mathbf{r}_{so}})$ pairs for all 40 ensemble members at the spatial point of the single observation. Two different kinds of datasets are shown here: first, the default dataset without an application of the TCI (squares) and, second, a dataset resulting from an application of the TCI (circles). Evidently, the TCI leads to a linear relationship between $(q_v^{\text{int}})^{(i)}_{\mathbf{r}_{so}}$ and $Z^{(i)}_{\mathbf{r}_{so}}$ (with nonzero slope) and, therefore, successfully increases the spread of the simulated reflectivities at the location of the single observation [Colour figure can be viewed at wileyonlinelibrary.com]

t_0 rejects any observed reflectivities in this region, no increments are produced, and, consequently, the LETKF is also not able to resolve the discrepancies between observed and simulated reflectivities; that is, the assimilation still lacks this particular observed cell.

We note that $\sigma[\mathbf{Z}]$ is only a linearized approximation here. As Z is not a model variable, it is not updated directly but implicitly through the update of other model variables. However, even though the dependence of Z on those other model variables is highly nonlinear and dynamic, the linearized increments of Z depending on $\sigma[\mathbf{Z}]$ represent a first approximation for how the model may respond to increments of model variables introduced via the LETKF.

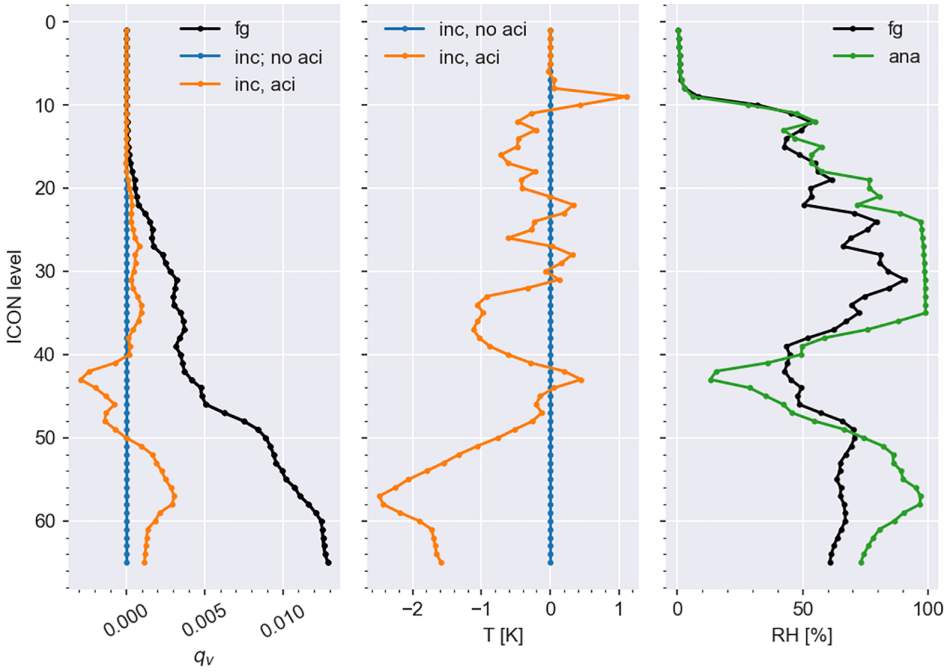

FIGURE 9 Vertical profiles of the (deterministic) first guesses, increments, and analyses for the specific humidity q_v (left), the temperature T (middle), and the relative humidity RH (right) at the horizontal position of the single observation at time t_0. The black lines indicate the values for the first guess at assimilation time t_0, the blue and orange lines indicate the values for the increment without and with an application of TCI, respectively, and the green line indicates the values for the analysis [Colour figure can be viewed at wileyonlinelibrary.com]

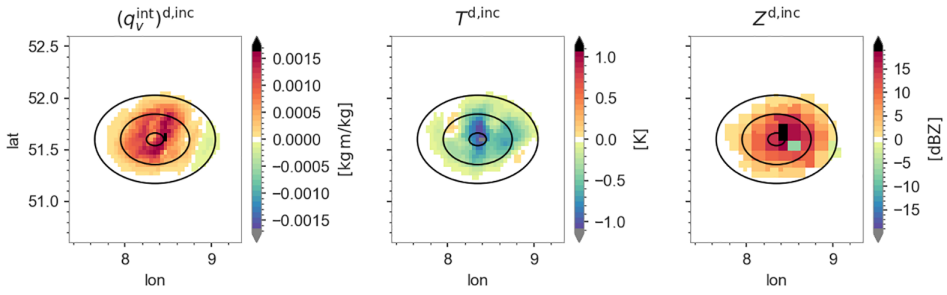

FIGURE 10 Increments for $(q_v^{int})^d$, T^d, and Z^d for the ICON layer 35 at t_0. The single observation position is indicated via concentric circles. It should be noted that the visualization for Z^d has a coarser resolution as it is given on a radar observation grid, while $(q_v^{int})^d$ and T^d are given on the model grid, which usually has a finer resolution [Colour figure can be viewed at wileyonlinelibrary.com]

Generally, for TCI to be applied at a specific location, it is reasonable to demand that several conditions be fulfilled. The first is that there is a zero spread for reflectivity. Furthermore, there should be a discrepancy between observed and simulated (deterministic) reflectivity, as we do not want to increase the spread at a location where there is not any reflectivity observed (and simulated) in the first place. For the final implementation of TCI, it includes a loop over all reflectivity observations which checks the above conditions, adds the TCI if they are satisfied, then proceeds with the LETKF as usual.

The location chosen for our single-observation experiments fulfills both of the requirements above, although we also made sure that we were selecting a location where even an assimilation of *all* radar observations does not lead to the production of a cell (that is otherwise observed) during the 60 min of a free forecast following the assimilation, such that the potential impact of TCI can be clearly seen.

4.2 | Applying the TCI

Now, we employ the TCI to inflate the spread of $\mathbf{Z}_{\mathbf{r}_{so}}$ based on the correlation of Z with q_v^{int}. A prerequisite for a successful application of the TCI is the existence of a sizeable spread of $(\mathbf{q}_\mathbf{v}^{\mathbf{int}})_{\mathbf{r}_{so}}$. This condition is fulfilled as demonstrated by Figure 7, where $(q_v^{int})^d$ and $\sigma[\mathbf{q}_\mathbf{v}^{\mathbf{int}}]$ are depicted.

Aiming for increments that are nearly maximized, we have to choose a suitable value for the observation error R and the TCI scaling factor α_{TCI}. Guided by the previously conducted statistical analysis from which we obtained $\alpha = 4.89\ \mathrm{dBZkgkg^{-1}\cdot m^{-1}}$, we use $\alpha_{TCI} = 5\ \mathrm{dBZkgkg^{-1}\cdot m^{-1}}$ as of now. Finding that $\sigma[(\mathbf{q}_\mathbf{v}^{\mathbf{int}})_{\mathbf{r}_{so}}] \approx 0.55\ \mathrm{kg\cdot m\cdot kg^{-1}}$, we obtain $\alpha_{TCI}\sigma[(\mathbf{q}_\mathbf{v}^{\mathbf{int}})_{\mathbf{r}_{so}}] \approx 2.75$ dBZ so that, with the help of Figure 4, we come to the choice of $R \equiv 4\,\mathrm{dBZ}^2$. This particular value for R has been chosen over the other alternatives as, firstly, $R = 4\,\mathrm{dBZ}^2$ is still a physically plausible value while producing a significantly increased increment compared with larger values for R and, secondly, it is still

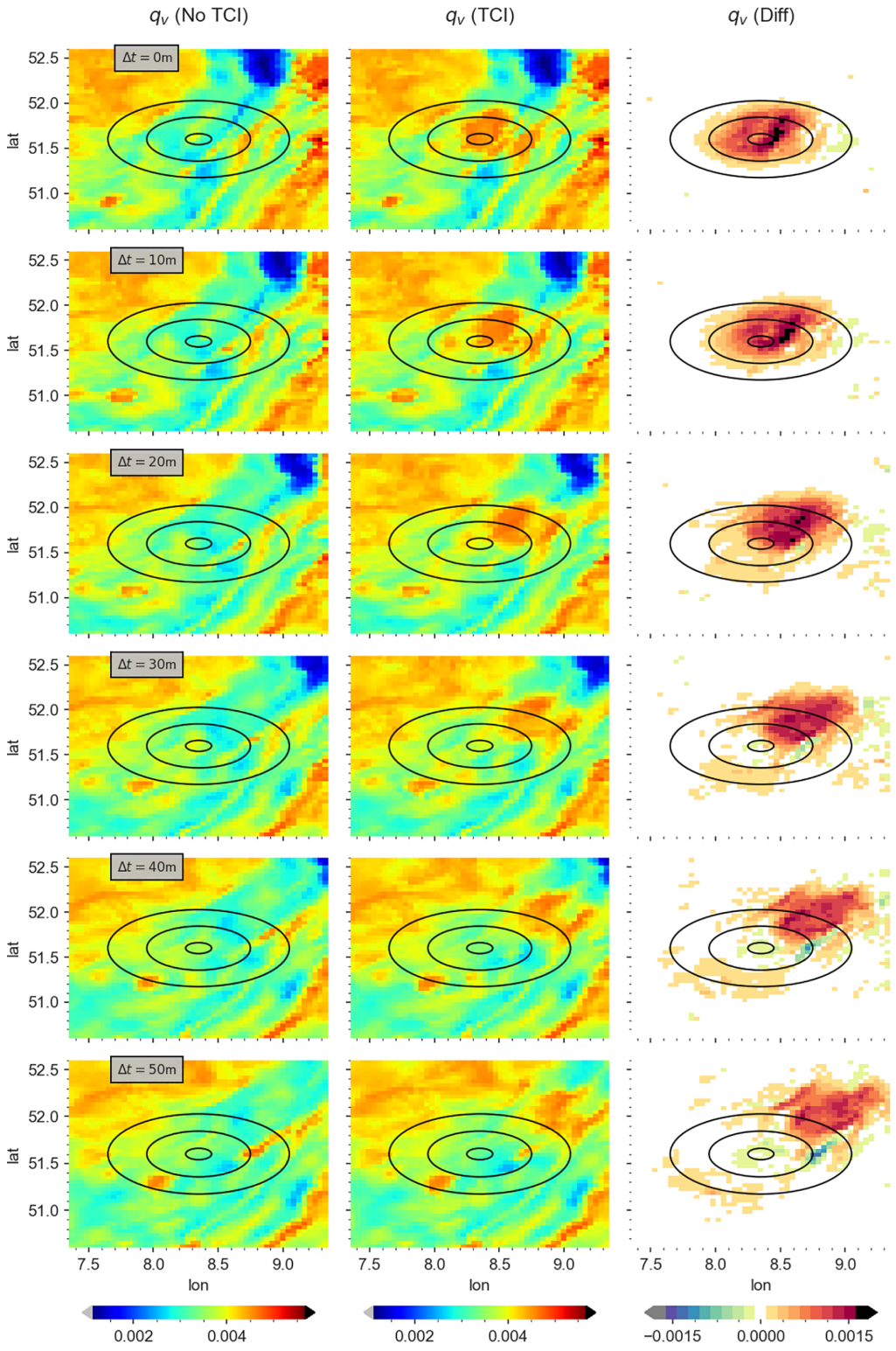

FIGURE 11 Humidity fields q_v at the ICON layer 35 without applying the TCI (first column), with applying the TCI (second column), and the difference of those two fields (third column) for forecast times of up to 50 min (row-wise) with respect to the assimilation date at t_0. The position of the single observation is indicated by concentric circles [Colour figure can be viewed at wileyonlinelibrary.com]

comparatively smooth in the region of $\alpha_{\mathrm{TCI}}\sigma[(\mathbf{q}_\mathbf{v}^{\mathbf{int}})_{\mathbf{r}_{so}}]$ as opposed to smaller values for R.

Figure 8 depicts pairs of $Z_{\mathbf{r}_{so}}^{(i)}$ and $(q_v^{\mathrm{int}})_{\mathbf{r}_{so}}^{(i)}$ for all ensemble members with and without applying the TCI, respectively. Without TCI, none of the ensemble members shows any reflectivity at $\mathbf{r}_{so}$, but the TCI leads to a linear relationship between $Z_{\mathbf{r}_{so}}^{(i)}$ and $(q_v^{\mathrm{int}})_{\mathbf{r}_{so}}^{(i)}$ with nonzero slope, as is to be expected when considering Equation (17).

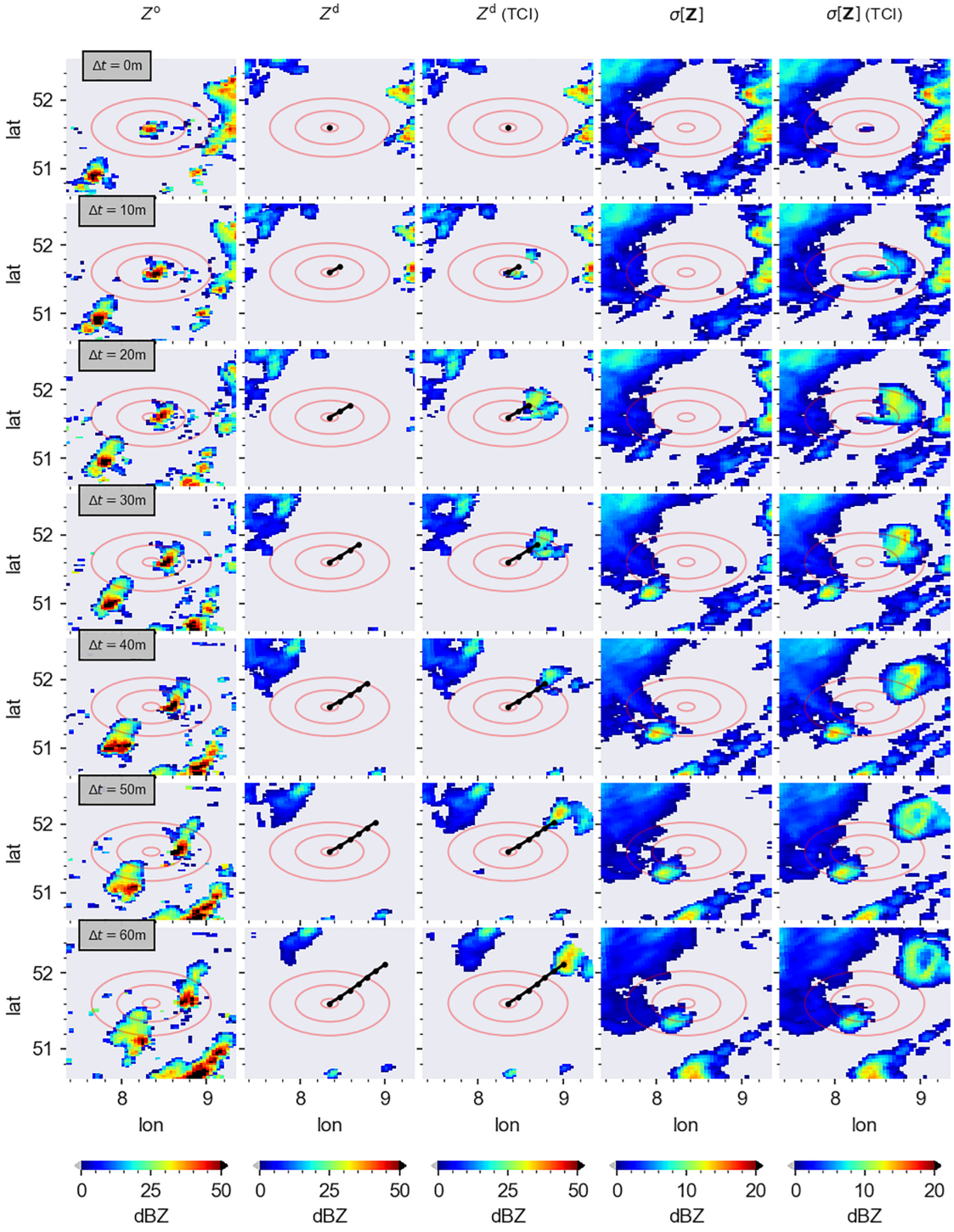

FIGURE 12 Radar composites for Z^o, Z^d, and $\sigma[\mathbf{Z}]$ for different forecast times of up to 60 min. Concentric circles indicate the position of the single observation. The black line indicates the trajectory obtained from a numerical integration of the equation of motion describing the movement of a particle starting at $\mathbf{r}_{so}$ within the external wind field [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | LETKF increments

An important model quantity involved in the production of reflectivity is the specific humidity q_v and the temperature T, whose increments are both shown in Figure 9. Without TCI, the increment is zero for both q_v and T, which is directly clear as the only assimilated observation is $Z^o_{\mathbf{r}_{so}}$ and $\sigma[\mathbf{Z}_{\mathbf{r}_{so}}]$ is vanishing. Applying the TCI, there are three pronounced peaks found for the increment of q_v around the ICON layers 35, 43, and 57 where the increment at layer 43 approximately amounts to the value of the background at this layer. Furthermore, the increments for

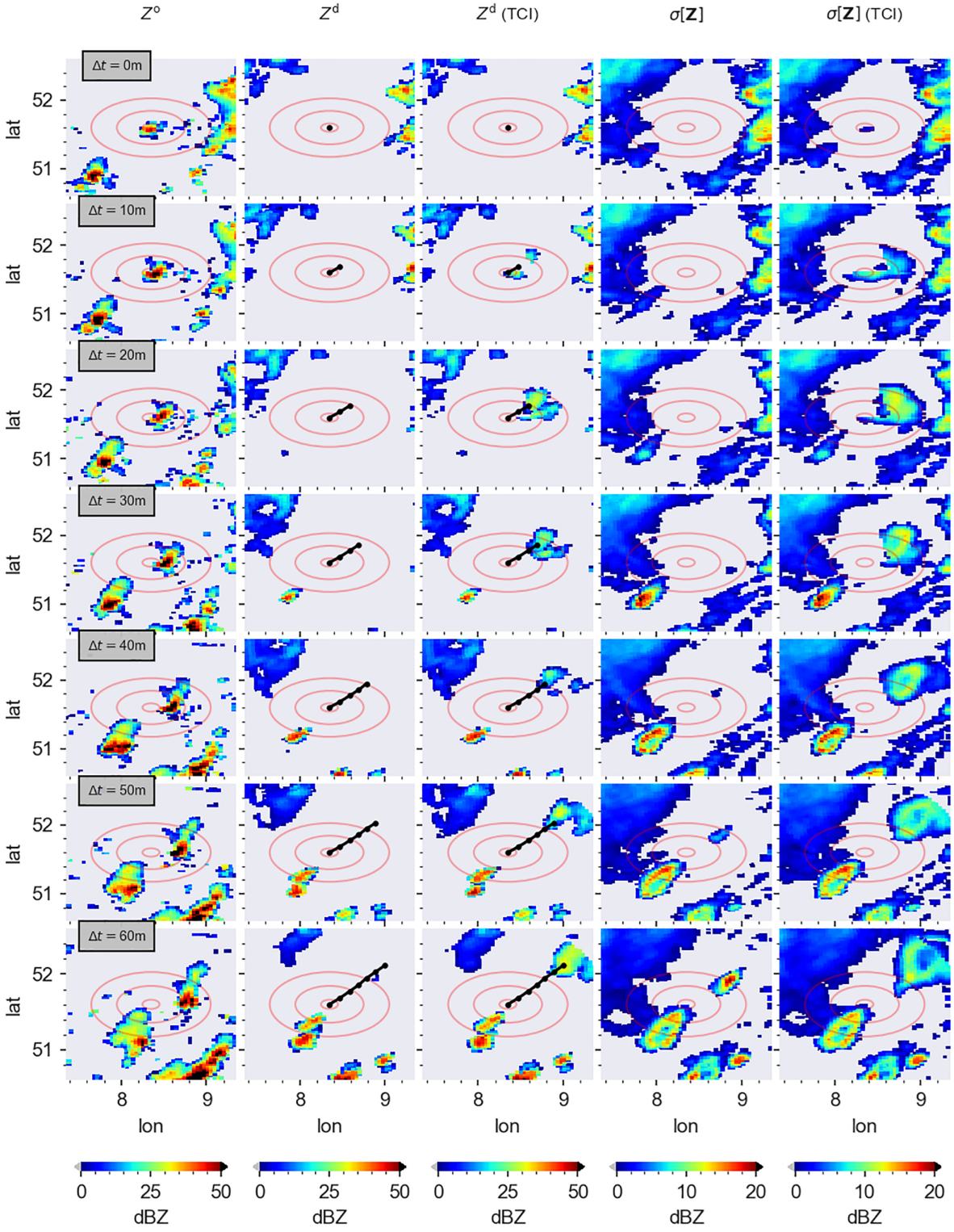

FIGURE 13 Similarly to Figure 12, this figure depicts radar composites for Z^o, Z^d, and $\sigma[\mathbf{Z}]$ for different forecast times of up to 60 min, but here LHN is also active. Concentric circles indicate the position of the single observation [Colour figure can be viewed at wileyonlinelibrary.com]

T are comparatively large around layer 57 and 45, where the largest increment is found in the vicinity of the ground, amounting to approximately 2.5 K. It is important to note that we use a large value for the vertical localization range such that the effect of vertical localization is negligible for our considerations here; that is, the increments imprinted on the complete vertical column are barely damped with increasing vertical distance from the single observation.

Figure 10 shows increments for q_v and T for a specific ICON layer and (linearized) increments for Z at the

elevation angle 0.5°. Here, the effect of the horizontal localization becomes directly visible: Nonzero increments are only found within a certain horizontal distance from the single-observation position—associated with the localization range—and become zero beyond this range.

4.4 | Impact on forecast

As the reflectivity is not a prognostic model variable, it is not updated directly but rather implicitly via the update of prognostic model variables such as q_v and T. Thus, we have to study how the model reacts to changes of increments of these prognostic variables that are eventually introduced by the TCI approach and investigate whether these changes lead to the new emergence of reflectivity.

The time evolution of q_v at the ICON layer 35 is depicted in Figure 11 starting at the assimilation time t_0 with free forecast times Δt of 1 hr for every 10 min. Evidently, the TCI leads to the appearance of additional humidity centered around the position of the single observation for $\Delta t = 0$. For $\Delta t > 0$, this additional humidity does not dissolve but is drifting in a northeast direction–which is consistent with the modeled wind fields in this region.

Similarly to Figure 11, Figure 12 shows the evolution of reflectivities with respect to forecast times Δt. The composites for Z^o show that the cell remains present for the complete forecast time of 60 min and that it is drifting in a northeast direction. The simulation without TCI, however, does not show any reflectivity for all forecast times, and as also the spread is zero within our region of interest, none of the ensemble members has any reflectivity there. The TCI approach and the additional increments for the humidity which it introduces into the model lead to the appearance of reflectivity starting at $\Delta t = 10$min. For $\Delta t > 10$min, the cell grows in horizontal area and, additionally, drifts in a northeast direction. The simulated cell seems to drift faster than the observed one, but the drift is consistent with a movement along the wind field predicted by the model.

4.5 | Comparison with LHN

Figure 13 shows, similarly to Figure 12, radar composites for observed and simulated reflectivities for several forecast times Δt. However, in contrast to our previous studies, LHN is also applied within the time frame of t_0 to $t_0 + \Delta t$. In this context, it is important to note that, in contrast to the TCI, LHN is applied to the complete domain and not only a single location. Furthermore, it should be noted that a double use of observations while simultaneously applying LHN and assimilating radar reflectivities is not an issue here since these two approaches rely on a separate execution of a precipitation scan and a volume scan, respectively, and, without going into too much detail here, employ rather different frameworks for further processing their observations. There are two particularly interesting observations that can be made here: Firstly, the LHN approach leads to the emergence of an additional simulated cell southwest of the single observation that gets increasingly larger over time and which is consistent with an observed cell in this region.

We note that this cell is only simulated with an application of LHN but not when only TCI is applied to a location. This seems clear since there is a large horizontal distance between this additional cell and the single observation. While we could have applied the TCI approach at a location near the origin of this additional cell, we wanted to demonstrate the principle of TCI in a very controlled setup in this work. Secondly, it is noteworthy that the LHN approach on its own, that is, without TCI, is not able to produce the cell that, considering the composites with an application of the TCI, first emerges in the vicinity of the single observation at $\Delta t = 10$min and then drifts in a northeast direction. This result indicates the significant potential of the TCI approach.

5 | SUMMARY AND OUTLOOK

In this work, we study a targeted covariance inflation (TCI) approach for improving the assimilation of 3D-volume radar reflectivities Z via the LETKF. While this approach formally belongs to the class of *additive* covariance inflation approaches, the background covariance matrix is not inflated by adding random perturbations, but the specific inflation is based on an empirical correlation of reflectivity with functions of model variables. We found vertically integrated specific humidity to be a promising functional candidate for adding artificial correlations by TCI.

Carrying out a case study, we investigate the effects of the TCI in a single-observation experiment, assimilating only a single reflectivity at a specific spatial position. This position was characterized by the fact that none of its ensemble members nor the deterministic run exhibit any reflectivity at assimilation time, although a sizeable reflectivity is observed. Without application of the TCI method, the increments obtained from the LETKF are zero due to the vanishing spread of Z. Applying the TCI, however, leads to a nonzero spread of Z and the LETKF now takes the observation into account and produces nonzero increments. These increments are generated based on the correlations with functions of specific humidity and added artificially, that is, outside of the LETKF formalism.

We evaluate TCI-based increments for temperature and specific humidity. Meteorological arguments show their potential to contribute to the production of convection and then reflectivity during a subsequent ICON model run. Our investigation shows that the model increment of specific humidity was "stable," that is, it did not disappear during a 1-hr model run following the assimilation. Furthermore, TCI successfully introduces a new convective cell in the direct vicinity of the assimilated single observation after a model evolution of 10 min. This newly created cell persists for the complete 60 min time window; it is consistent with the observed cell, and drifts along the wind fields in a northeastern direction. Furthermore, within our case study, we compare the effect and capability of the TCI approach with LHN, observing that LHN alone is not able to generate the desired cell. Though only being a case study at this point, this demonstrates the successful implementation and the potential of targeted covariance inflation.

Clearly, much further work needs to be done to fully develop TCI based on the results of this work. A more systematic study within the full convective-scale model domain and more extended simulation periods is necessary. However, we also need to investigate the applicability of targeted covariance inflation to other observations of clouds, for example, visible reflectances measured by satellites. When the model does not generate clouds in a given column, LETKF faces similar challenges as for radar observations and TCI could be an approach to help resolve the problems which LETKF faces in this situation.

AUTHOR CONTRIBUTIONS

Klaus Vobig: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review and editing. **Klaus Stephan:** conceptualization; methodology; supervision; writing – review and editing. **Ulrich Blahak:** conceptualization; methodology; supervision; writing – review and editing. **Kobra Khosravian:** **Roland Potthast:** conceptualization; funding acquisition; methodology; project administration; supervision; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

Open Access funding enabled and organized by Projekt DEAL.

ORCID

Klaus Vobig https://orcid.org/0000-0002-6399-3527

REFERENCES

Anderson, B.D.O. and Moore, J.B. (2012) *Optimal Filtering. Dover Books on Electrical Engineering Series*. New York, NY: Dover Publications, Incorporated.

Anderson, J. and Anderson, S.L. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127, 2741–2758.

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

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, 139(12), 3887–3905.

Bannister, R.N. (2017) A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 607–633.

Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y. and Potthast, R. (2016) Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1490–1504.

Dowell, D.C. and Wicker, L.J. (2009) Additive noise for storm-scale ensemble data assimilation. *Journal of Atmospheric and Oceanic Technology*, 26(5), 911–927.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), 10143–10162.

Evensen, G. (2009) *Data Assimilation: The Ensemble Kalman Filter. Earth and Environmental Science*. Heidelberg, Germany: Springer.

Evensen, G. and van Leeuwen, P.J. (2000) An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128, 1852–1867.

Houtekamer, P.L. and Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796–811.

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

Houtekamer, P.L. and Mitchell, H.L. (2005) Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3269–3289.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L. and Hansen, B. (2005) Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Monthly Weather Review*, 133(3), 604–620.

Houtekamer, P.L. and Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12), 4489–4532.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2006) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *arXiv*. arXiv: physics/0511236.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2), 112–126.

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

Kleist, D.T., Parrish, D.F., Derber, J.C., Treadon, R., Wu, W.-S. and Lord, S. (2009) Introduction of the GSI into the NCEP global data assimilation system. *Weather and Forecasting*, 24(6), 1691–1705.

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J. and Saunders, F.W. (2000) The Met Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570), 2991–3012.

Nakamura, G. and Potthast, R. (2015) *Inverse Modeling*, (pp. 2053–2563). Bristol, England: IOP Publishing.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D. and Yorke, J.A. (2004) A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56(5), 415–428.

Prill, F., Reinert, D., Rieger, D. and Zaengl, G. (2020). ICON Model Tutorial 2020. DWD.

Reich, S. and Cotter, C. (2015) *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge, England: Cambridge University Press.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA): ensemble data assimilation for the COSMO Model. *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1453–1472.

Snyder, C. and Zhang, F. (2003) Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, 131(8), 1663–1677.

Stephan, K., Klink, S. and Schraff, C. (2008) Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quarterly Journal of the Royal Meteorological Society*, 134(634), 1315–1326.

van Leeuwen, P.J., Cheng, Y. and Reich, S. (2015) *Nonlinear Data Assimilation. Frontiers in Applied Dynamical Systems: Reviews and Tutorials*. Heidelberg, Germany: Springer.

Whitaker, J.S. and Hamill, T.M. (2002) Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 1913–1924.

Yokota, S., Seko, H., Kunii, M., Yamauchi, H. and Sato, E. (2018) Improving short-term rainfall forecasts by assimilating weather radar reflectivity using additive ensemble perturbations. *Journal of Geophysical Research: Atmospheres*, 123(17), 9047–9062.

Zeng, Y., Blahak, U. and Jerger, D. (2016) An efficient modular volume-scanning radar forward operator for NWP models: description and coupling to the COSMO model: an Efficient Radar Forward Operator and Coupling to the COSMO Model. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 3234–3256.

Zhang, F., Snyder, C. and Sun, J. (2004) Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132, 1238–1253.

Zängl, G., Reinert, D., Rípodas, P. and Baldauf, M. (2015) The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 563–579.

How to cite this article: Vobig, K., Stephan, K., Blahak, U., Khosravian, K. & Potthast, R. (2021) Targeted covariance inflation for 3D-volume radar reflectivity assimilation with the LETKF. *Quarterly Journal of the Royal Meteorological Society*, 147(740), 3789–3805. Available from: https://doi.org/10.1002/qj.4157