

A composite state method for ensemble data assimilation with multiple limited-area models

By MATTHEW KRETSCHMER^{1*}, BRIAN R. HUNT², EDWARD OTT^{1,3}, CRAIG H. BISHOP⁴, SABRINA RAINWATER⁵ and ISTVAN SZUNYOGH⁶, ¹*Department of Physics, University of Maryland, College Park, MD, USA;* ²*Department of Mathematics, University of Maryland, College Park, MD, USA;* ³*Department of Electrical Engineering, University of Maryland, College Park, MD, USA;* ⁴*Naval Research Laboratory, Monterey, California, CA, USA;* ⁵*National Research Council, Monterey, CA, USA;* ⁶*Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA*

(Manuscript received 31 October 2014; in final form 17 March 2015)

ABSTRACT

Limited-area models (LAMs) allow high-resolution forecasts to be made for geographic regions of interest when resources are limited. Typically, boundary conditions for these models are provided through one-way boundary coupling from a coarser resolution global model. Here, data assimilation is considered in a situation in which a global model supplies boundary conditions to multiple LAMs. The data assimilation method presented combines information from all of the models to construct a single ‘composite state’, on which data assimilation is subsequently performed. The analysis composite state is then used to form the initial conditions of the global model and all of the LAMs for the next forecast cycle. The method is tested by using numerical experiments with simple, chaotic models. The results of the experiments show that there is a clear forecast benefit to allowing LAM states to influence one another during the analysis. In addition, adding LAM information at analysis time has a strong positive impact on global model forecast performance, even at points not covered by the LAMs.

Keywords: Ensemble Kalman Filter, limited-area models, composite state

1. Introduction

Geophysical fluid dynamical forecasts (e.g. for atmospheric or oceanic states) are typically created by integrating a numerical model forward in time from suitable initial conditions. Limited-area models (LAMs), which only cover a restricted geographic area, allow high-resolution forecasts for small, subglobal regions of interest to be made when limited resources prevent running a high-resolution global forecast model or assimilating high-resolution global data. While higher spatial resolution does not guarantee that a model will behave more realistically, it is generally helpful in this endeavour. (For simplicity, our numerical experiments are designed to ensure that increased model resolution corresponds to increased model realism.)

LAMs are found in a variety of settings, ranging from research to operational situations. Typically, operational weather centres run LAMs that are defined over the geo-

graphic region(s) of interest to that centre. Some centres, such as those of the U.S. National Weather Service and the U.S. Navy, run their LAM independently on multiple domains that may or may not overlap, to produce high-resolution regional forecasts. The U.S. Navy, in particular, routinely runs their limited-area COAMPS[®] model on more than 70 different limited-area domains (Pielke, 2013). Some of these domains are overlapping, and their union covers more than 20% of the globe. But, in the data assimilation phase, the analysis state of one LAM is normally not directly affected by the state of any of the others, and the analysis state of the global model providing LAM boundary conditions is not affected by the analysis LAM states. The present work investigates the possibility of improving forecast performance by allowing interactions between global and limited-area model states at analysis times in situations where multiple LAMs are employed.

In particular, we consider data assimilation in an ensemble forecasting context. Ensemble data assimilation uses a collection of model forecasts to estimate the probability distribution of the state of the system. We investigate an approach in which each individual LAM is used to integrate

*Corresponding author.
email: mkretsch@umd.edu

its own ensemble of model states. Integrating an ensemble of LAM model states requires an ensemble of boundary conditions that transmit information about the large scale flow features to each of the LAM ensemble members. An ensemble of lateral boundary conditions may be generated in a number of ways (Torn et al., 2006), which includes concurrently running a global forecast ensemble (Merkova et al., 2011; Holt et al., 2013).

Traditionally, data assimilation has been performed on the global and limited-area models separately. However, recently it has been shown that using the global analysis (Guidard and Fischer, 2008) or a short-range forecast based on the global analysis (Dahlgren and Gustafsson, 2012) as an additional constraint on the limited-area state estimate has a positive effect on the limited-area analysis, while allowing communication between the global and limited-area processes can improve both the global and the limited-area analyses (Yoon et al., 2012). This finding is similar to the one that two-way grid nesting during the forecasting phase can help reduce discrepancies between global and limited-area model states (Harris and Durran, 2010). In situations with multiple LAMs, two-way coupling of the models during the analysis procedure can potentially improve the global analysis. More accurate global analyses would, in turn, lead to improvements in the lateral boundary conditions applied to LAMs. Allowing for additional communication between LAMs during the analysis may help alleviate the lateral boundary condition errors that have typically plagued LAMs (Warner et al., 1997).

Here we present a framework that combines all available global and limited-area model state information at analysis time, and performs data assimilation on this composite state. An aim of the new approach is to obtain ensemble mean analyses for all LAMs and the global model that minimise the distance between the analysed ensemble mean and truth filtered to the resolution that each model is supposed to simulate. Obtaining analyses that lie on the model attractors is not an explicit part of this aim; but to the extent that the model is like filtered reality, obtaining analyses close to filtered reality will provide initial conditions close to the model attractors. Recent work by Klocke and Rodwell (2014) suggests that even the most advanced current data assimilation systems provide analyses that are pushed off the model attractor in the direction of the true filtered state. The authors of that paper also showed that the mean short-term initial drift of the model forecast from the analysis toward the model attractor provides highly useful information for the identification and correction of model errors.

Yoon et al. (2012) used a *joint-state* approach in which the components of the state vector analysed by the Local Ensemble Transform Kalman Filter (LETKF) (Ott et al., 2003; Hunt et al., 2007) consisted of the components of the regional and global state vectors. A drawback of

this approach was that an *ad hoc* term had to be added to the state estimation equations to moderate a tendency of the global and regional state estimates to diverge.

The new approach presented here is based on the notion that there is just one true state and that the object of data assimilation should be to estimate that state as accurately as possible. At locations where LAMs are present, it is possible to estimate the true state at a higher resolution than at locations where LAMs are not present. Hence, in the new composite state method introduced in the present paper, the LETKF is applied to a single, variable resolution state vector in which the local resolution is determined either by the highest resolution LAM available at that location or by the global model. The forecasts for this variable resolution state vector are obtained by linearly combining information from models that have grid boxes that overlap with points on the variable resolution state vector. Applying the LETKF to this variable resolution state vector results in an ensemble of variable resolution analyses. These variable resolution analyses can then be filtered down to the lower resolution models that cover the same area. In this way, the composite state approach delivers analyses to both the global model and the regional models that are informed by all the models.

A theoretically appealing aspect of the composite state approach is that it would deliver the minimum error variance state estimate provided the correct forecast and observation error covariance matrices were specified and the effects of localisation in the LETKF were negligible. In contrast, the presence of the *ad hoc* term in the formulation of Yoon et al. (2012) precludes the possibility of it delivering the minimum error variance state estimate.

We test our method in a series of simulated observation experiments that utilises the simple, one-dimensional, chaotic models introduced by Lorenz (2005), and we find that analyses and forecasts produced by a coarse resolution ‘global’ model and a collection of high-resolution ‘LAMs’ that cover the entire simulation domain can attain essentially the same accuracy as analyses and forecasts produced by a single high-resolution global model. If only part of the global domain is covered by overlapping LAMs, the analyses and forecasts are essentially as accurate as if the overlapping LAMs were replaced with a single LAM covering the same region, and except near the boundary of this region, the results can be nearly as accurate as using the high-resolution model globally. These results serve as motivation for further investigation, to see if real systems utilising our method might realise similar forecast improvements. Importantly, unlike the Lorenz models, real-world atmospheric models include interaction between motions at a wide range of scales. Further studies will be necessary to explore the behaviour of these scale interactions in model simulations that use the composite state method.

The rest of this paper is organised as follows. Section 2 introduces the composite state technique. Section 3 describes our numerical experiments with the Lorenz (2005) models. We use a sparse observation network, so that the advantage of the high-resolution LAMs over the global model is primarily in greater forecast accuracy, rather than an ability to resolve the observations. In Section 4, we consider the case where the LAMs cover the global domain. In this case, the composite state analysis is done on a global high-resolution grid, and any disadvantage compared to using a single global high-resolution model must be due to the decomposition of the global model states into LAM states for the forecast phase of the analysis cycle. We explore the effect of varying LAM region size and global domain size within this context. We find that there is essentially no degradation of the composite state estimate relative to using a single high-resolution model, unless the LAM region size is too small. In Section 5, we consider overlapping LAMs that cover only part of the global domain, so that the analysis is done at coarse resolution on the part of the globe that is not covered by LAMs. As in Section 4, the interface between neighbouring LAMs is not a significant source of error; degradation compared with using a global high-resolution model occurs near the boundary with the coarse resolution region. Section 6 presents our conclusions and further discussion.

Finally, we note that our presentation is in the context of one-dimensional models defined on subsets of the same common grid. In practice, one would be interested in three-dimensional atmospheric models and a collection of LAMs, each with its own grid. These issues are not directly addressed here, but would have to be dealt with if our method were to be operationally applied (see Section 6).

2. Data assimilation and the composite state method

Data assimilation is a cyclic alternation between a short-term forecast and a procedure, called the ‘analysis’, that seeks to estimate a system’s state (and possibly its error statistics). At the beginning of each cycle, the analysis procedure is performed, combining available observational information with a forward forecast from the end of the previous cycle (called the ‘background’ estimate) to yield an updated estimate of the system state (the ‘analysis state’). The estimates of the uncertainties in the observations and background state estimate are crucial to this step, as they determine the relative weighting of the observations and background in forming the analysis state estimate. Once the analysis procedure is completed, the analysis state, and possibly its uncertainty, can be forecast to the next analysis time, where the cycle is repeated.

Data assimilation based on Ensemble Kalman Filters (Evensen, 1994; Burgers et al., 1998), which will be the basis of our work, has attracted much recent interest because of its ability to form consistent time-dependent uncertainty estimates in the analysis. In an Ensemble Kalman Filter, a collection of system state estimates (the ensemble) is evolved in time and updated with observational information at the start of each analysis. The best guess of the system state is given by the ensemble mean, and the background error covariance matrix P_b , a measure of the state estimate’s uncertainty, is approximated by the sample covariance of the background ensemble. Here, we conduct our analysis on the composite state ensemble using the LETKF (Hunt et al., 2007). The LETKF algorithm seeks an analysis ensemble with mean and covariance given by the Kalman Filter update equations. It is equivalent to a localised version of the ETKF of Bishop et al. (2001) with spherical simplex centring (Wang et al., 2004), and may be viewed as a computationally advantageous version of the LEKF method of Ott et al. (2004). A local ETKF has been successfully implemented to generate ensemble perturbations for an operational global-regional model pair in the UK Met Office’s MOGREPS system (Bowler and Mylne, 2009). The LETKF has also been tested, with positive results, using the regional models of the Italian and German weather services (Bonavita et al., 2010; Reich et al., 2011; Lange and Craig, 2014).

In Ensemble Kalman Filters, the background and analysis ensembles are typically expressed as an ensemble mean, and a collection of ensemble perturbations from this mean. The ensemble mean is denoted $\bar{\mathbf{x}}$, an N vector, with N being the dimension of the model state. The ensemble perturbations form an $N \times k$ matrix \mathbf{X} , where k is the number of ensemble members. The m th column of this matrix ($m = 1, 2, \dots, k$) represents the perturbation ($\mathbf{x}_m - \bar{\mathbf{x}}$) of the m th ensemble member, \mathbf{x}_m , from the ensemble mean. In the LETKF, the analysis ensemble mean and ensemble perturbations are expressed in terms of the background ensemble mean and ensemble perturbations through a weight vector, \mathbf{w} , and a weight matrix, \mathbf{W} ,

$$\begin{aligned}\bar{\mathbf{x}}^{(a)} &= \bar{\mathbf{x}}^{(b)} + \mathbf{X}^{(b)}\mathbf{w}, \\ \mathbf{X}^{(a)} &= \mathbf{X}^{(b)}\mathbf{W}.\end{aligned}\tag{1}$$

Here $\mathbf{X}^{(b)}$ and $\mathbf{X}^{(a)}$ are the matrices of background and analysis ensemble perturbations, respectively. The background and analysis ensemble means are given by $\bar{\mathbf{x}}^{(b)}$ and $\bar{\mathbf{x}}^{(a)}$, respectively. The LETKF conducts its analysis in a local ensemble space. At each grid point n , the LETKF uses the observations within an empirically determined local analysis region to determine \mathbf{w} and \mathbf{W} for grid point n , and hence the analysis state value at that grid point.

Although our presentation and numerical examples assume use of the LETKF framework, our method does

not rely on the details of the LETKF, and we expect it can be straightforwardly adapted to other versions of Ensemble Kalman Filters (see, e.g., Houtekamer and Mitchell, 1998; Anderson and Anderson, 1999; Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002).

2.1. The composite state method

In the following discussion, we consider cases with a global model, whose state is denoted \mathbf{x}_0 , and c LAMs, with states labelled \mathbf{x}_i , where $i=1,2,\dots,c$. Our method does data assimilation on all model ensemble states simultaneously by performing the analysis on a ‘composite’ state ensemble. The composite state is defined on a model grid with potentially non-uniform spatial resolution that, in a given region, matches that of the highest resolution model defined there. More formally, denoting the region where the i th LAM is defined as L_i , the entire global model domain as L_0 , and a location in the continuous global domain as A , the composite lattice has the same resolution at A as the global model when A is not in $\bigcup_{i=1}^c L_i$. Here the notation $\bigcup_{i=1}^c L_i$ denotes the subset of the global domain covered by all LAMs. When A is in $\bigcup_{i=1}^c L_i$, the composite state lattice has the same resolution at A as the highest resolution LAM defined there.

We view the state estimates contained in each forecast model ensemble to be approximations to the same true state, albeit with different accuracies. We consider situations in which high spatial-resolution short-term forecasts will be more accurate than low spatial-resolution short-term forecasts, and thus we value those forecasts produced at higher spatial resolution more when constructing the composite state ensemble. This assumption, and the algorithmic simplifications that it permits for performing data assimilation, is what led us to the specific formulation of the composite state method presented here. In formulating our composite state method, we envision a situation in which there are no persistent forecast model biases in the LAM and global model forecasts, so that the dominant forecast errors inside the LAM domain are caused by spatial truncation errors. (The subject of correcting forecast model biases will be a subject of future work.) The m th ensemble member of the composite state ensemble, $\hat{\mathbf{x}}_m$, is constructed from the m th ensemble members of the global and each LAM ensemble. For brevity, in the following discussion we suppress the ensemble member subscript, m . The composite state $\hat{\mathbf{x}}(n)$ at location n is a linear function of the state vectors of the global and limited-area models whose domains contain n , as given by eq. (2),

$$\hat{\mathbf{x}}(n) = \sum_{i=0}^c p_i(n) O_n[\mathbf{x}_i]. \quad (2)$$

The functions $p_i(n)$ define the weighting given to the i th LAM model state at location n . The operator O_n interpolates a state vector \mathbf{x}_i to location n when n does not correspond to a grid point of \mathbf{x}_i , and acts as the identity operator in situations when n corresponds to a grid point of \mathbf{x}_i . If all LAMs whose domains contain a location n have the same spatial resolution, but incommensurate grids, O_n will still need to interpolate, but the choice of which LAM grid to interpolate to is left to the user. For all $i=0,1,\dots,c$, the functions $p_i(n)$ satisfy $0 \leq p_i(n) \leq 1$ and $\sum_{i=0}^c p_i(n) = 1$ at every location n , and $p_i(n) = 0$ if n is outside the domain L_i of the i th LAM. In general, $p_i(n)$ should vary slowly with n , to ensure continuity of $\hat{\mathbf{x}}(n)$. However, if the LAM is evolved with boundary conditions from the global model, so that $\mathbf{x}_i = \mathbf{x}_0$ on the boundary of L_i , then $p_i(n)$ and $p_0(n)$ can change discontinuously as n crosses the boundary. Within these restrictions, there is a lot of freedom in picking the values of the $p_i(n)$ functions when LAMs overlap. In Section 3 we specify reasonable $p_i(n)$ choices for the situations we consider in our experiments.

The first step in performing data assimilation using the composite state method is to create a background ensemble of composite state vectors from the background ensembles of global and limited-area model states, using the definition of the composite state, eq. (2). Once the background composite state ensemble has been created, the analysis procedure is carried out on it, yielding the analysis composite state ensemble. The global and limited-area model state ensembles that are forecasted to the next analysis time are each constructed from the analysis composite state ensemble, interpolating from the composite state grid to the appropriate global or limited-area lattice when necessary. In the LETKF formulation, the interpolation is best done on the weight vector \mathbf{w} and matrix \mathbf{W} , applying the interpolated weights to the model states on their native grid (Yang et al., 2009).

3. Numerical experiments

For ease of presentation, in the following sections we specialise to the context of one spatial dimension, evenly spaced grid points forming a high-resolution grid, and a low-resolution grid of evenly spaced grid points that are a subset of this high-resolution grid. The global model is taken to be defined on the low-resolution grid, and each LAM is defined on a subset interval of the high-resolution grid. For this situation, the operator O_n in eq. (2) always acts as the identity, and can thus be omitted. In addition, we consider the case where at most two LAM domains can overlap at any given location. At locations where only the i th LAM is defined, $p_i(n) = 1$, and outside of the domain of the i th LAM, $p_i(n) = 0$. In our experiments, we chose $p_0(n) = 0$ at all locations n where LAMs are defined, and

$p_0(n) = 1$ at locations where only the global model is defined. If LAM i overlaps with another LAM, we chose $p_i(n)$ to decrease linearly across the overlap region, to zero at the edge of LAM i . We chose this form of weighting function for its simplicity, and because of our observation that boundary condition errors decreased with increasing distance from the LAM's lateral boundary (see Fig. 1). For other models, it may be advantageous to have the global weights $p_0(n)$ taper more continuously from 1 to 0 near the boundary of the LAM domains, or remain positive throughout the LAM domains. Additionally, our experiments use LAMs governed by the same model physics, at the same spatial resolution. More complex scenarios may benefit by choosing the $p_i(n)$ functions to more heavily weight LAMs with better historical error properties. As an illustrative case for the example of two LAMs with domains as shown in Fig. 2a, Fig. 2b–d show corresponding choices of $p_0(n)$, $p_1(n)$ and $p_2(n)$.

3.1. The Lorenz models

To test our data assimilation framework we perform a series of numerical experiments. These tests require three models: a global high-resolution model which generates the

simulated ‘nature’ or ‘truth’; a high-resolution LAM model; and a low-resolution global model that has only large spatial-scale behaviour.

With these considerations in mind, we utilise in our experiments two models that are described in Lorenz (2005), known as Lorenz Models II and III. Lorenz Model II describes the spatiotemporal behaviour of a quantity, Z , defined on a one-dimensional lattice with periodic boundary conditions, analogous to a ring of constant latitude. A subscript n is used to index the value of Z at each grid point on this lattice. Model II exhibits spatially extended and smooth waves, and is given by

$$\frac{dZ_n}{dt} = [Z, Z]_{K,n} - Z_n + F. \quad (3)$$

The terms on the right hand side (RHS) of eq. (3) are analogous to nonlinear advection, dissipation, and forcing, respectively. The bracket term in eq. (3) is given by

$$[X, Y]_{K,n} = \sum_{j=-J}^J \sum_{l=-J}^J (-X_{n-2K-l} Y_{n-K-j} + X_{n-K+j-l} Y_{n+K+j}) / K^2. \quad (4)$$

The forcing parameter, F , and smoothing parameter, K , in eq. (3) and eq. (4) are chosen by the experimenter, with

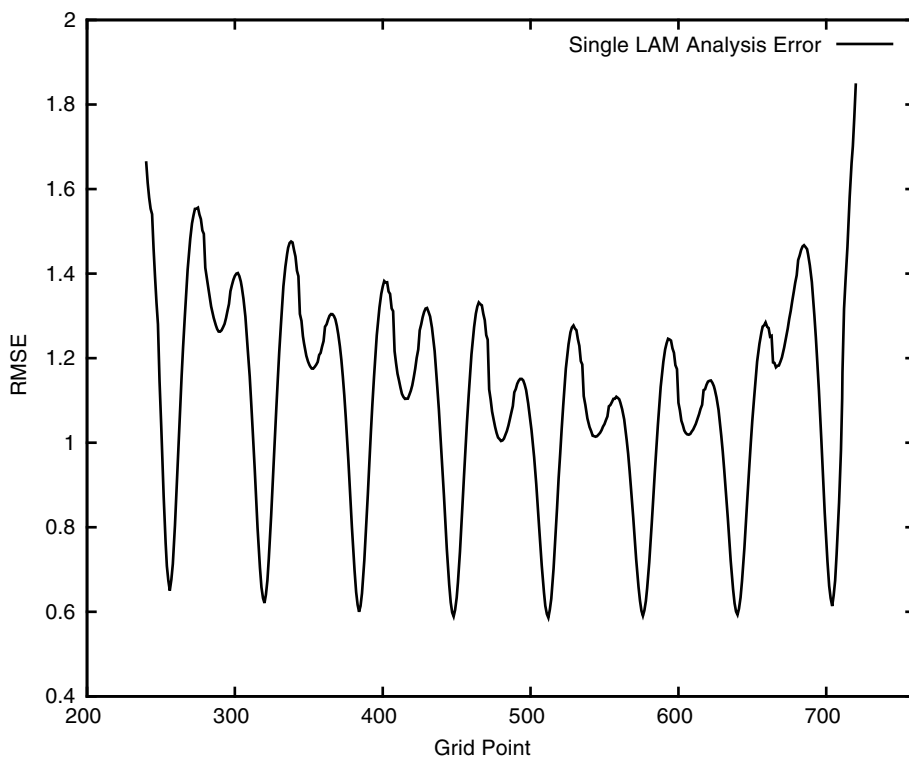


Fig. 1. RMS analysis errors of the ensemble mean when the LETKF is performed using only the state information of a single LAM. The LAM is defined over grid points [240,720]. The RMS errors shown are averaged over 2×10^4 analysis cycles. Boundary condition errors can be seen in the increase in RMS error at grid points near the LAM boundary.

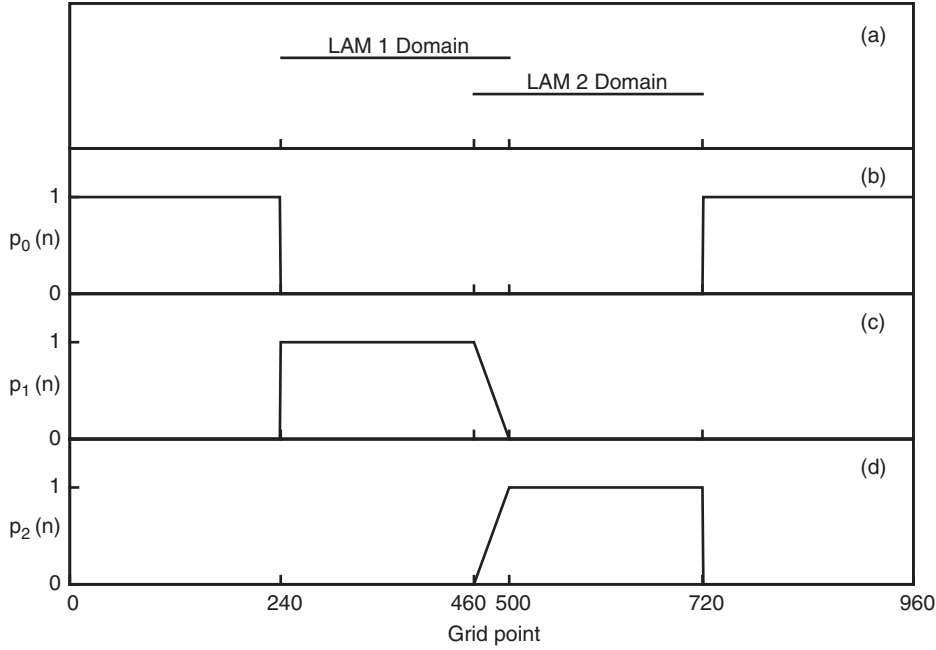


Fig. 2. An example of the p_i functions for a scenario with two limited-area models, denoted LAM 1 and LAM 2, used in one of our experiments. Part (a) shows the domains on which the LAMs are defined, which cover grid point intervals of [240,500] and [460,720]. Parts (b)–(d) show the functional form of the p_i functions for the global model, $p_0(n)$ [shown in panel (b)], and each of the limited-area models, $p_1(n)$ and $p_2(n)$ [shown in panels (c) and (d), respectively]. All plots show grid point location n on the horizontal axis.

$J = K/2$ when K is even, and $J = (K-1)/2$ when K is odd. The primed summation, \sum' , represents a modified summation, with the first and last terms divided by 2 when K is even. If K is odd, \sum' denotes a normal summation.

As noted in Lorenz (2005), the waves of Model II have long spatial scales, making it fitting to represent a coarse ‘global’ model. Smaller spatial-scale behaviour can be added to the dynamics of the quantity Z by modifying the equations of Model II, to arrive at Model III,

$$\frac{dZ_n}{dt} = [X, X]_{K,n} + b^2[Y, Y]_{1,n} + c[Y, X]_{1,n} - X_n - bY_n + F. \quad (5)$$

In Model III, the quantity Z varies spatially with long and short spatial-scale components, denoted by X and Y , respectively. The first two terms on the RHS of eq. (5) represent spatially smoothed nonlinear advection of the long spatial-scale component X and nonlinear advection of the short spatial-scale component Y , respectively. The long and short scale components are coupled through the third term on the RHS of eq. (5). The last terms on the RHS of eq. (5) represent linear damping and constant forcing. The parameters b and c control the amplitude of short scale waves and the coupling between scales, respectively. We use Model III to represent both the LAM and ‘nature’

dynamics. The long spatial-scale component of a Model III state is found at grid point n via spatial averaging of Z ,

$$X_n = \sum_{l=-I}^I (\alpha - \beta|l|) Z_{n+l},$$

and the short scale component is found by $Y_n = Z_n - X_n$. The summation limit I is chosen by the experimenter. The α and β quantities are functions of I , whose exact form can be found in Lorenz (2005). These are chosen so that $X_n = Z_n$ whenever Z_n varies quadratically within $\pm I$ grid points of n . Like Model II, Model III is also defined on a one-dimensional lattice with periodic boundary conditions, although typically at a higher spatial resolution. As noted by Lorenz, one model time unit of these models is analogous to approximately 5 d in the atmosphere.

3.2. Experimental parameter and domain details

In our experiments, we use Model III with parameter values of $K=32$, $b=10$, $c=0.6$, $F=15$, and $I=12$ to govern the LAM and ‘nature’ model dynamics. The global model is Model II, with parameters $F=15$ and $K=8$, and $1/4$ the spatial resolution of the ‘nature’ model. In both Models II and III, K is a smoothing parameter which controls the spatial resolution of the long wavelength waves. (Additional experiments used $c=2.5$ in the LAM and nature model

dynamics, enhancing the difference between the global and limited-area and nature model dynamics. Despite more dramatic differences between the global and limited-area model attractors, the composite state method performed similarly to the experiments reported on here.) The ‘nature’ model lies on a lattice of $n=960$ grid points, which are indexed from $n=0, \dots, 959$. The LAM domains are not constant from one experiment to the next, but are always defined on continuous subsets of the ‘nature’ grid, such as $n=0, 1, \dots, 540$ or $n=480, 481, \dots, 60$, for example. In most of our experiments, we consider a global model defined on a lattice of 240 grid points, each corresponding to every 4th ‘nature’ model grid point, $n=0, 4, 8, 12, \dots, 956$. In experiments where the global model resolution is lowered by a factor of two, K is adjusted to $K=4$. To avoid ambiguity, we index all model grid points by their corresponding ‘nature’ grid point index.

The global and limited-area model ensembles have 40 ensemble members in our experiments. The initial Model II global ensemble is sampled from a free run of the global model, after 600 model ‘days’ of spin-up time. At the beginning of each experiment, each LAM ensemble member is produced by interpolating a corresponding initial global Model II ensemble member onto the finer LAM grid, at locations where the LAM is defined. The initial conditions of the ‘nature’ model are found by initialising its grid points with random numbers uniformly distributed in the interval $[0, 1]$, and allowing the model to spin up for 600 model days.

During our numerical experiments, we perform data assimilation every $dt=0.05$ model time units [the equivalent of about every 6 hours in real time, according to Lorenz (2005)]. At each analysis time, observations of the ‘nature’ model at 15 equally spaced observation locations, located at grid points $0, 64, \dots, 896$, are created by adding Gaussian noise with mean 0 and standard deviation 1 to the ‘nature’ model values. A key parameter of the LETKF is the size of the local state vector patch on which the analysis is performed. For our experiments, we use a local patch size that is 81 grid points wide, so that at least one observation is assimilated at every grid point. The influence of observations inside of a local analysis region was not tapered for the experiments reported on here; we expect that results could improve when that technique is implemented in the LETKF. Also, multiplicative covariance inflation is used in our experiments to prevent filter divergence (Anderson and Anderson, 1999). At each analysis cycle, the composite ensemble background covariance matrix is inflated by the constant factor $\rho=1.048$, which was found through empirical tuning to minimise RMS error.

3.3. Numerical integration

The global and regional ensembles are forecasted simultaneously, using a fourth order Runge-Kutta scheme, breaking the ‘6-hour’ forecast between analysis times into 36 time steps. The boundary conditions required by each LAM ensemble member are provided by its corresponding global ensemble member. Specifically, interpolation of the state values of the appropriate global ensemble member is used to provide state values needed in eq. (4) that are outside of the LAM domain, but which are needed for the integration of eq. (5).

During the integration, we utilise Davies Relaxation (Davies, 1983). We define ‘sponge regions’ at both boundaries of each LAM domain, having a length of 10 LAM grid points. After the ensembles have been integrated forward in time by 1 time step, the state of an ensemble member of LAM i at a grid point in a sponge region is updated to a linear combination of the corresponding forecasted LAM i and global model ensemble members at that grid point, according to

$$\mathbf{x}_i(n) \rightarrow (1 - \gamma(n))\mathbf{x}_i(n) + \gamma(n)\mathbf{x}_0(n). \quad (6)$$

Here, $\mathbf{x}_i(n)$ and $\mathbf{x}_0(n)$ are the state values at grid point n of a member of limited-area ensemble i and a global model ensemble member, respectively, and $\gamma(n)$ is a spatially dependent weighting function. In our experiments, $\gamma(n)$ decreases linearly over the sponge region, from a value of 1 at the outer sponge region boundary to 0 at the inner sponge region boundary. At grid points in the sponge region at which the global model is undefined, the global state value is linearly interpolated onto the finer LAM mesh, and this value is used for $\mathbf{x}_0(n)$ in eq. (6).

3.4. Verification details

The results presented below use the temporally averaged Root-Mean-Square Error (RMSE) between the ensemble mean, $\bar{\mathbf{x}}$, and the truth \mathbf{x}_t

$$\text{RMSE}(n) = \sqrt{\frac{1}{T} \sum_{q=1}^T (\bar{\mathbf{x}}_q(n) - \mathbf{x}_{t,q}(n))^2},$$

as a measure of the effectiveness of our method. Here the subscript q indexes analysis cycle. The RMSE of the ensemble mean at grid point n is the average, over T analysis cycles, of the squared difference between the ensemble mean and the truth at n , square-rooted. Errors are calculated at each analysis time, as well as for single deterministic forecasts. These forecasts use analysis ensemble means as initial conditions.

We test the composite state method in two situations, one where the LAMs collectively cover the global domain,

$L_0 = \bigcup_{i=1}^c L_i$, and one where they do not. In each case, we generate forecasts, initialised from the analysis ensemble mean, for both the global and limited-area models.

We compare results using the composite state method to high- and low-resolution forecasts made from a pair of models that are defined over the entire experimental domain. These benchmark high- and low-resolution ensemble forecasts are created using the same high- and low-resolution models used to create the composite state method forecasts, Lorenz Models III and II, respectively. The benchmark ensemble forecasts assimilate the same set of observations as the composite state forecasts, using the same algorithm, the LETKF, with the same number of ensemble members. The benchmark ensembles are integrated for 6 model hours between analysis cycles. The composite state analysis ensemble mean is verified against the benchmark analysis ensemble mean after each analysis. For forecasts longer than 6 hours, both benchmark and composite state forecast estimates are found by integrating the benchmark and composite state analysis ensemble means, using the appropriate forecast models.

For the experiments in which the LAMs do not collectively cover the entire experimental domain, we compare composite state analyses and forecasts to those from the ‘joint-state method’ of Yoon et al. (2012). The joint-state method performs data assimilation simultaneously on both global and limited-area models. It accomplishes this by using an observation function that predicts the observations by using information from both the global model state and the LAM state, as well as by including a constraint term in the cost function that penalises large differences between global and limited-area model states. We compare the high-resolution limited-area forecasts of the joint-state method to similar high-resolution forecasts created using the composite state method.

Our method differs from the joint-state method in a number of ways, one of which is that the joint-state method does not constrain the global and limited-area model analyses to be identical. More specifically, the joint-state method performs data assimilation by minimising a local cost function, $J^n(\mathbf{x})$, at each grid point. This local cost function depends upon the local background ensemble, local observations, and their respective error covariances, and is given by

$$J^n(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_b)^T \mathbf{P}_b^{-1} (\mathbf{x} - \bar{\mathbf{x}}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) + \kappa (\mathbf{x}_g - \mathbf{x}_r)^T (\mathbf{x}_g - \mathbf{x}_r). \quad (7)$$

Here $\bar{\mathbf{x}}_b$ is the local background ensemble mean, \mathbf{P}_b is the local ensemble sample covariance, \mathbf{y} is a vector of local observations, \mathbf{R} is the local observation error covariance matrix, and \mathbf{x}_g and \mathbf{x}_r are vectors which contain

the global and regional (LAM) model state values, respectively, at grid points at which both the global and regional ensembles are defined. The κ in eq. (7) is a parameter. Importantly, the forward operator H used in the joint-state method depends on location. Inside of the local analysis region, it maps a linear combination of global and regional model states to observation space, and depends upon a parameter λ . Both of these parameters, λ and κ , must be tuned for optimal performance in any application of the joint-state method. In applications with d limited-area models, the joint-state method would necessitate the empirical tuning of $2d$ of these parameters, in addition to parameters associated with other empirical techniques, such as covariance inflation. Additionally, the computational cost of finding the minimum of eq. (7) grows quickly with the number of LAMs present in a local analysis region, as each of the d LAMs would contribute a term equivalent to the third term in eq. (7). Both of these qualities make implementing the joint-state method in any context with multiple LAMs exceedingly complicated. The composite state method proposed here presents a simpler approach to performing data assimilation in the multiple LAM context that is a natural extension of the joint-state method, avoids the necessity of empirically tuning a large number of adjustable parameters, and allows a considerably simpler cost function to be minimised. The composite state method corresponds to the case where $\lambda = 1$ and $\kappa = \infty$ in eq. (7). For additional details on the joint-state method, see Yoon et al. (2012).

As a diagnostic of the composite state method, we conducted ensemble forecasts, verified at several lead times, to measure the relationship between ensemble spread and ensemble RMS error while using the composite state method, and found that the global model ensemble spread adjusts appropriately to match the decreased RMS ensemble forecast error of the composite state method. Globally averaging over space and time, the ensemble spread for 6-hour forecasts was found, for the cases considered, to be approximately equal to the ensemble RMS error. Specifically, these quantities differed by approximately 2%.

4. Results for global LAM coverage

As a first test of the composite state method, we apply it to a situation where there are two LAMs, whose domains together cover the entire ‘global’ domain of our experiments. Both LAMs are driven at their boundaries by the global model dynamics. When there is no communication between global and limited-area model ensembles during the analysis it is typical that boundary conditions supplied this way lead to an increase in errors near LAM boundaries, as a result of mismatches in state information at these locations. To see how such errors can be eliminated by the

composite state method, Fig. 1 provides an illustrative example of lateral boundary-induced LAM errors, for the case of a single LAM defined over the interval $[240,720]$ of our $n=0$ to 960 grid, when assimilation is performed separately on the limited-area and global model states, with no feedback between limited-area and global model state information during the analysis. While both ensembles of model states assimilate the same observations, the ensemble of LAM states is used as the background ensemble only for the LAM assimilation, and the ensemble of global states is used as the background ensemble only for the global assimilation. Figure 1 shows how error near both LAM boundaries can be dramatically larger than the error closer to the LAM domain interior.

In contrast to the separate analysis method shown in Fig. 1, we find that applying the composite state method in multiple LAM situations allows LAM analysis and 1-d forecast errors to rival those of a globally high-resolution perfect model, as demonstrated in Figs. 3 and 4a, for a situation with two overlapping LAMs that cover the entire $n=0$ to 960 domain. For the case shown in Figs. 3 and 4a, the LAMs are defined over the intervals $[0,520]$ and $[480,960]$, overlapping at 41 grid points near each of their boundaries. Similar LAM analysis and forecast accuracies

were achieved with other overlap values, as shown in Fig. 5. In the experiments whose results are shown in Fig. 5, analysis accuracy of the composite state analysis ensemble mean is calculated when two equally sized LAMs, whose domains collectively cover the entire experimental domain, are used to construct the composite state ensemble. The analysis RMS error is calculated as a function of LAM domain overlap, with larger overlap corresponding to larger LAM domain size. As the size of the LAM domain overlap grows, we see from Fig. 5 that there is virtually no change in analysis or forecast errors, indicating that there is little benefit to large LAM domain overlap.

In Figs. 3 and 4a, we can see that the RMS analysis and 1-d forecast errors obtained using the composite state method (red and green curves, respectively) are virtually the same as those obtained using a global high-resolution perfect model (black curve). Figure 4b shows 5-d forecast statistics, produced by limited-area and global models initialised to the composite state analysis ensemble mean. For comparison, curves of the RMS error of 5-d forecasts produced by global low- and high-resolution ensembles without the composite state method are also shown. Comparing the errors of composite state (blue curve) and control (orange curve) low-resolution global model 5-d

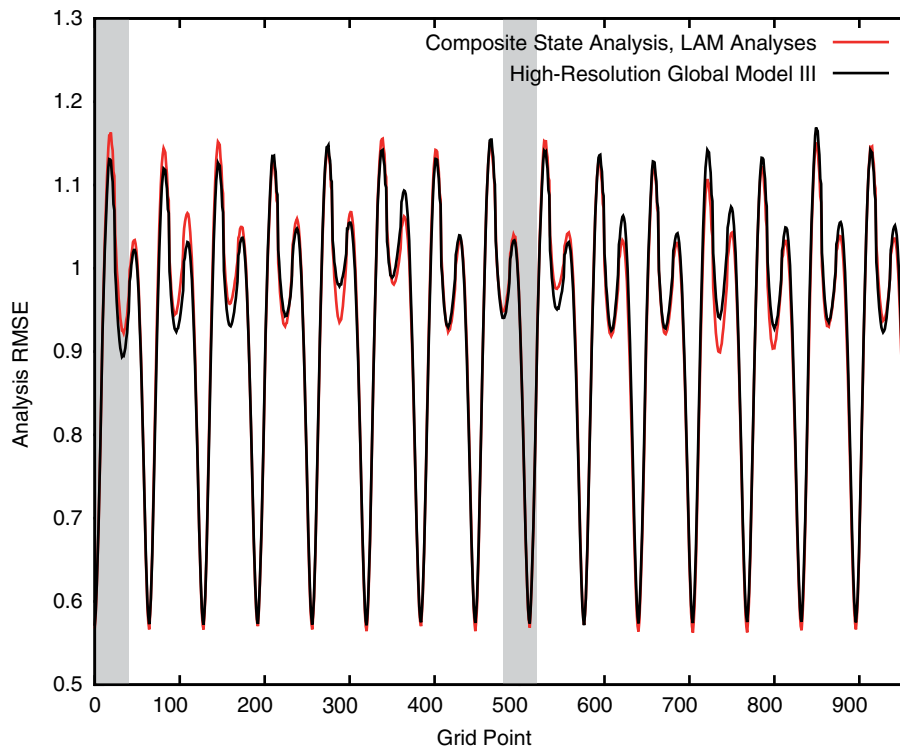


Fig. 3. RMS analysis errors of the composite state ensemble mean (red curve). For comparison, the analysis error of the ensemble mean for a global high-resolution perfect model LETKF analysis (black curve) is also shown. LAMs are defined over grid points $[0,520]$ and $[480,960]$, and statistics are taken over 10^5 analysis cycles, discarding the first 10^3 cycles. The shaded areas indicate the domain where both LAMs are defined.

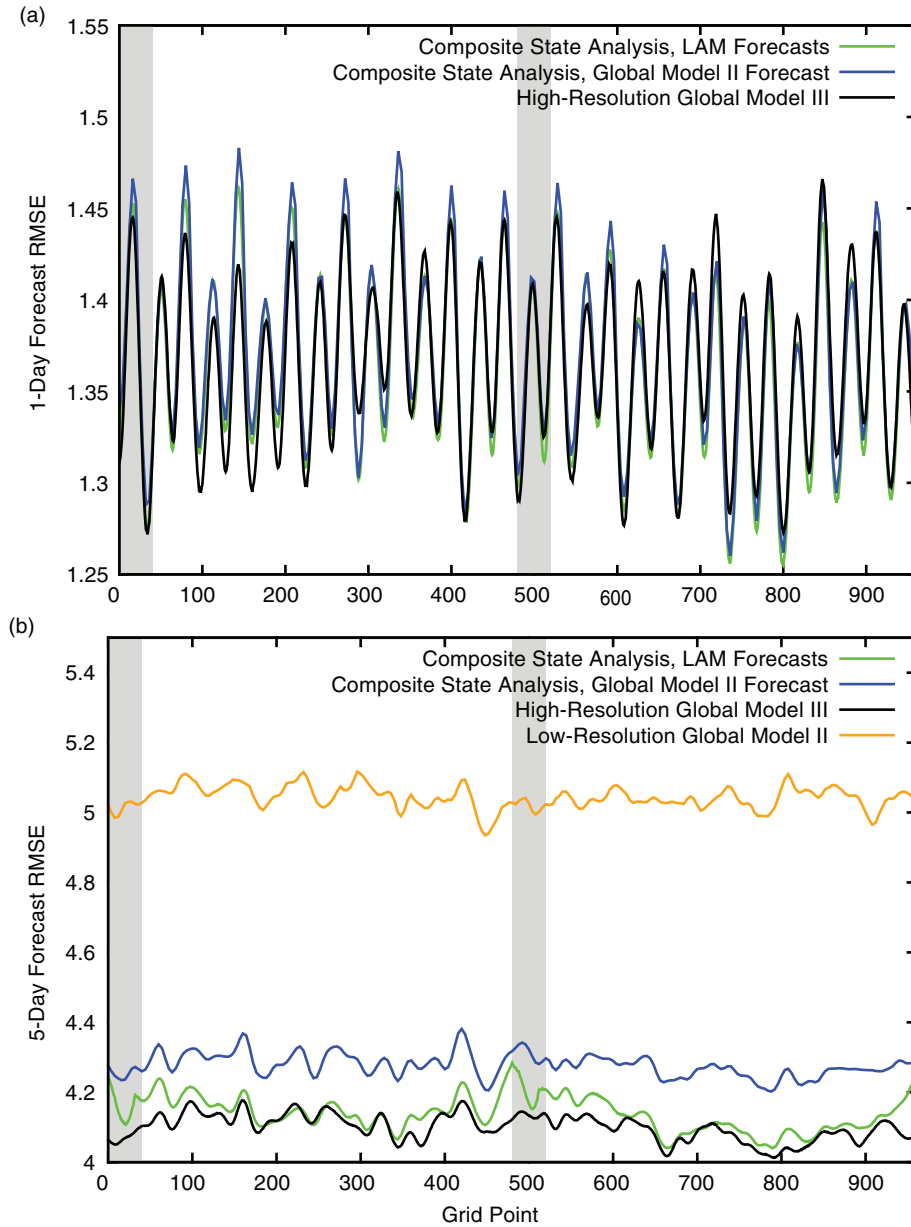


Fig. 4. (a) RMS 1-d and (b) 5-d forecast errors, initialised using the composite state analysis ensemble mean. The green curves in both panels show errors of forecasts produced by the LAMs, while the blue curves show errors for forecasts produced by the low-resolution, imperfect global model. For comparison, forecast errors produced by a global high-resolution perfect model and a global low-resolution imperfect model, initialised from an LETKF analysis are shown as black curves [in panels (a) and (b)] and an orange curve [in panel (b)], respectively. Both panels show results using the conditions described in Fig. 3, and the shaded areas indicate regions of LAM domain overlap.

forecasts, the composite state method substantially improves the global model 5-d forecasts, indicating that much of the difference between the high- and low-resolution global model forecasts (black and orange curves, respectively) is due to initial condition errors.

The LAMs are able to approach the accuracy of high-resolution global model forecasts (black curve in Fig. 4b) in

the interior of the LAM domains. The effects of imperfect boundary information coming from the global model can be seen in Fig. 4b, as LAM forecast errors rise near LAM boundaries. The size of the effected region is dictated by the flow of imperfect state information into the LAM from the lower resolution global model. As shown by Yoon et al. (2010), this information moves predominantly ‘eastward’

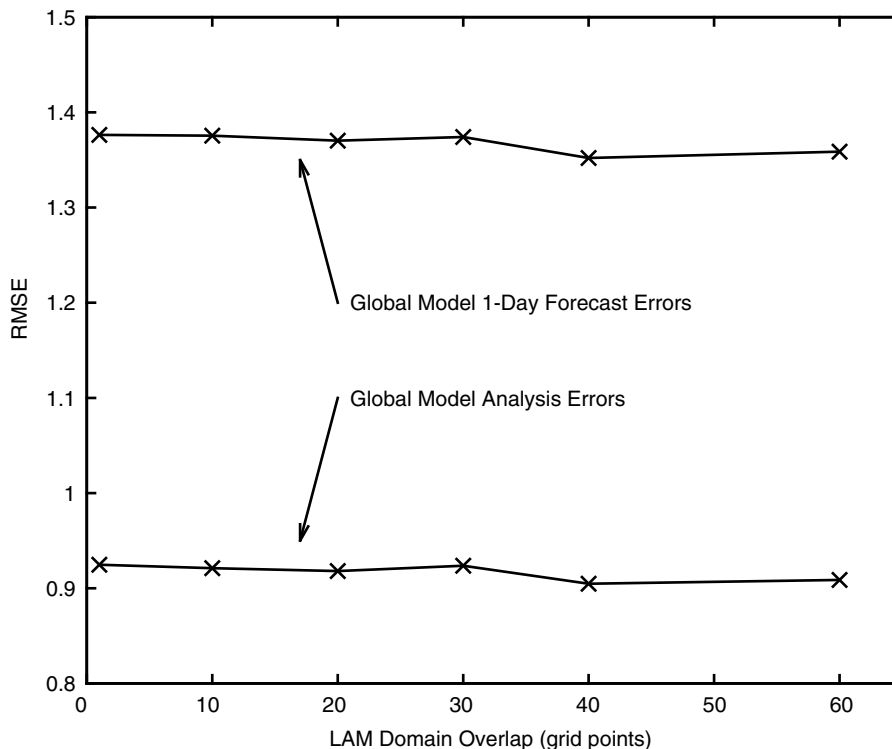


Fig. 5. RMS analysis and 1-d forecast errors of the global model ensemble mean, averaged over all grid points and time (10^5 analysis cycles), discarding 10^3 initial spin-up cycles, and performing the analysis using the composite state method. The results shown above are for two LAMs whose domains tile the globe. The x-axis shows the number of grid points that the LAM domains have in common. For these models, there appears to be no benefit to large LAM domain overlap.

(direction of increasing n), at a rate of 1.4 grid points per ‘hour’ in the Lorenz models. Thus for a 5-d (120-hour) forecast, on the order of 160 grid points will be adversely affected by boundary condition errors, which is approximately what can be seen in Fig. 4b, in the grid point intervals $[0,160]$ and $[480,640]$. Also, compatible with the predominantly ‘eastward’ propagation of information in the Lorenz models (Yoon et al., 2010), we see that the adversely affected region of the LAM domain is larger to the ‘east’ of a LAM boundary than to the ‘west’ of a boundary. Choosing LAM domains with greater overlap can help mitigate the effect of LAM boundary errors on forecasts, as with sufficient overlap grid points affected by boundary conditions in one LAM might correspond to more accurate, interior grid points of another.

To see how the global model resolution influences the accuracy of these results, we apply the composite state method in an experiment with two LAMs defined over the intervals $[0,540]$ and $[480,60]$, but do so using a global model with $K=4$ that is defined on every eight of the nature model grid points, or half of the resolution of the previously used global models. To quantify how much the RMS forecast and analysis errors change when the global model resolution is lowered, we compare spatiotemporally

averaged RMS errors of composite state analyses and 1-d forecasts made using global models at both of these spatial resolutions. The lowering of resolution causes the RMSE of the composite state analysis mean to increase by approximately 2.9%, and the 1-d forecast RMSE of the LAMs to increase by approximately 1.7%, while decreasing global forecasting costs by 50%, as a result of decreased function evaluations. Thus, for our original setup, we see that the resolution of the global model may be lowered without much loss of accuracy.

We now test the composite state method on multiple LAMs defined over a larger experimental domain that consists of twice as many grid points (1920) as in the experiments just described, but maintains spatial resolution and Model II and III parameter values (e.g. $K=32$ for LAM and nature dynamics). Thus, there are now four LAMs in this ‘Large World’ experiment, each defined over 541 grid points. The observation density is also held constant. The 1-d forecast results from this experiment are shown in Fig. 6. The composite state method is again seen to achieve performance that is virtually the same as that of a global high-resolution perfect model. In this scenario, the errors are approximately constant across LAM domains and there are no large deviations in RMSE at the

LAM boundaries (shaded regions) similar to those seen at the LAM boundaries in Fig. 1, indicating that lateral boundary errors have been minimised. The results of this experiment lead us to believe that the composite state method is scalable with LAM number, and similar forecast accuracies may be achieved with a much larger number of LAMs, in a proportionately much larger global domain.

Another condition we investigate is the size of LAM domains. In order to test this we consider an experiment on our original $n=0$ to 960 grid that uses an increasing number of LAMs. Specifically, we calculate errors of the composite state ensemble mean when two, four, eight, and 16 identically sized LAMs are independently forecasted during each analysis cycle; results of these experiments are shown in Fig. 7. These LAMs have sizes of 521, 261, 131, and 65 grid points, with overlaps of 41, 21, 11, and 5 grid points, respectively. The concurrently running global model ensemble provides boundary conditions to each of these individual LAMs. Despite the smaller domain sizes of the experiments with four or eight LAMs, Fig. 7 shows that there is not much loss of accuracy with these domain sizes. In Fig. 7 we can see that errors begin to increase when the LAM size is small enough such that on the forecast phase of the analysis cycle, boundary errors from the driving

global dynamics are able to affect a larger proportion of the LAM domain. This increase in error is not because of smaller LAM overlap, as Fig. 5 shows that forecast accuracy is not strongly dependent on LAM domain overlap. As the speed of information flow in the Lorenz models is approximately 1.4 grid points per hour (Yoon et al., 2010), in one 6-hour analysis cycle information travels about 10 grid points. For a 24-hour forecast time, information from the boundaries would affect LAM grid points up to approximately 30 grid points inside the LAM domain, and it is unsurprising that LAM domains large enough such that these boundary regions represent a small fraction of overall size would exhibit similar forecast results.

5. Results for incomplete LAM coverage of global domain

As we have seen, analyses and forecasts produced using the composite state method can rival those produced by a high-resolution ensemble of perfect model states when there is a collection of LAMs which cover the entire experimental ‘globe’. We now show that forecasting systems comprised of a single limited-area and global model pair can realise

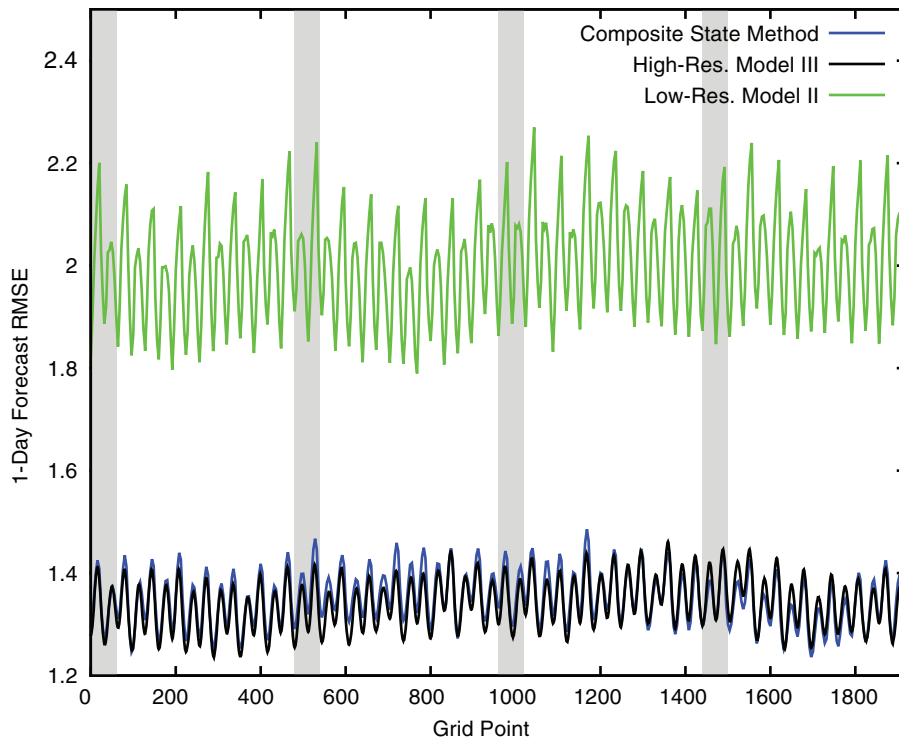


Fig. 6. RMS 1-d forecast errors, averaged over 10^5 analysis cycles, in the ‘Large World’ scenario. The experimental domain runs from $n=0$ to 1920, and LAMs are defined over grid point intervals $[0,540]$, $[480,1020]$, $[960,1500]$ and $[1440,60]$. The blue curve shows forecasts, initialised using the composite state analysis ensemble mean, made with the low-resolution global model. Observations are located at every 64 grid points, and the shaded areas indicate grid point intervals where more than one LAM domain is defined.

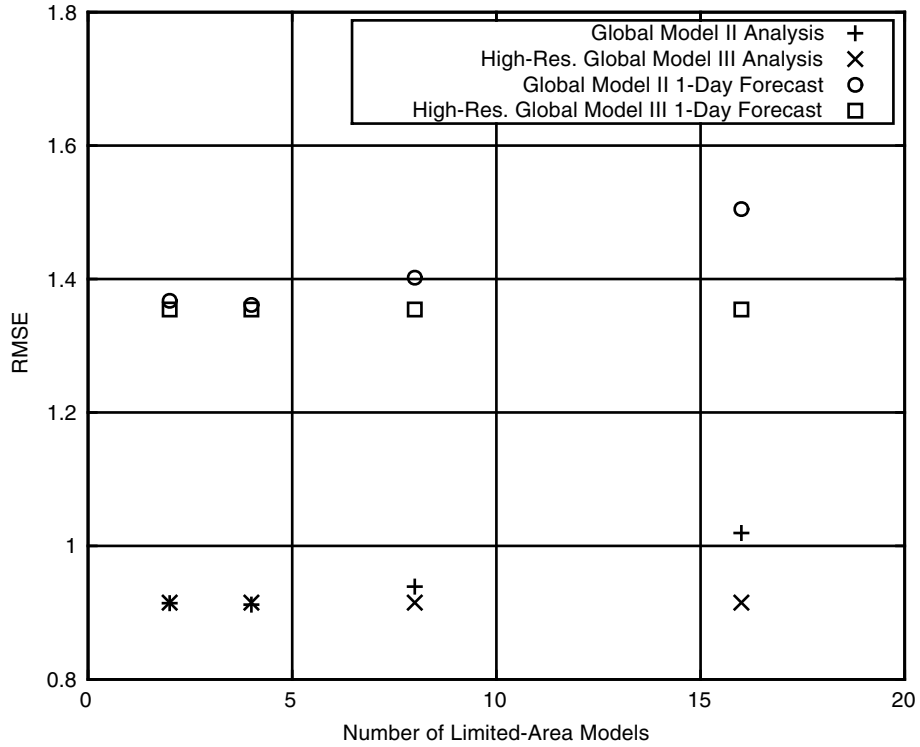


Fig. 7. RMS analysis errors of the ensemble mean and 1-d forecast errors calculated using the composite state method, when the entire simulation domain is divided amongst different numbers of LAM domains. In a given experiment, each of the LAM domains are the same size, so that the LAM domains are 521 grid points long in the two LAM case, 261 grid points long in the four LAM case, 131 grid points long in the eight LAM case, and 65 grid points long in the 16 LAM case. Errors begin to increase as the area influenced by boundary condition errors becomes a larger part of the total LAM domain. Statistics are averaged first over 10^5 analysis cycles, discarding the first 10^3 cycles, then over all grid points.

dramatic benefits to analysis and forecast accuracy if state information from an additional LAM is included in the analysis procedure. To do this, we first find analysis errors for a single limited-area and global model pair, when the analysis is performed using both the composite state method and the joint-state method of Yoon et al. (2012). These errors are then compared to those calculated using the composite state method when there are two LAMs and a global model. In this second situation, the LAMs are defined over the domains [240,720] and [720,240], collectively covering the globe, but with overlap only at the two boundary points.

Figure 8a shows the results of these experiments as curves of the RMS error of the analysis composite state ensemble mean, for the LAM defined over the interval [240,720] of a $n=0$ to 960 grid point domain. This result demonstrates that for the single LAM case, the composite state analysis error (blue curve) is approximately the same as that calculated using the joint-state method (brown curve). However, when we add another LAM to the analysis the accuracy of the composite state method analysis (green curve in Fig. 8a) becomes competitive with the high-resolution global perfect

model ensemble (black curve), as shown above in Section 4 (see Fig. 3). The previous strong increase of analysis error near the left LAM boundary is nearly eliminated when this extra LAM state information is considered in the analysis procedure. (The asymmetry between left and right boundaries is a result of the eastward (direction of increasing n) ‘group velocity’ of waves in the Lorenz models, which is analogous to the westerly atmospheric flow in the Northern Hemisphere mid-latitudes.) Comparing the blue curve in Fig. 8a to the curve in Fig. 1 shows that the composite state method helps alleviate increases in boundary error near the right LAM boundary as well.

Even more dramatic effects are seen in Fig. 8b, which shows the benefits to 1-d forecast accuracy when state information from an additional LAM ensemble is considered in the analysis. The composite state method considers this extra information in the analysis and can produce an analysis ensemble whose mean allows more accurate LAM and global model forecasts than would otherwise be possible. We note that the addition of a LAM allows the composite state method to greatly decrease the forecast error near the left-most (‘western’) LAM boundary.

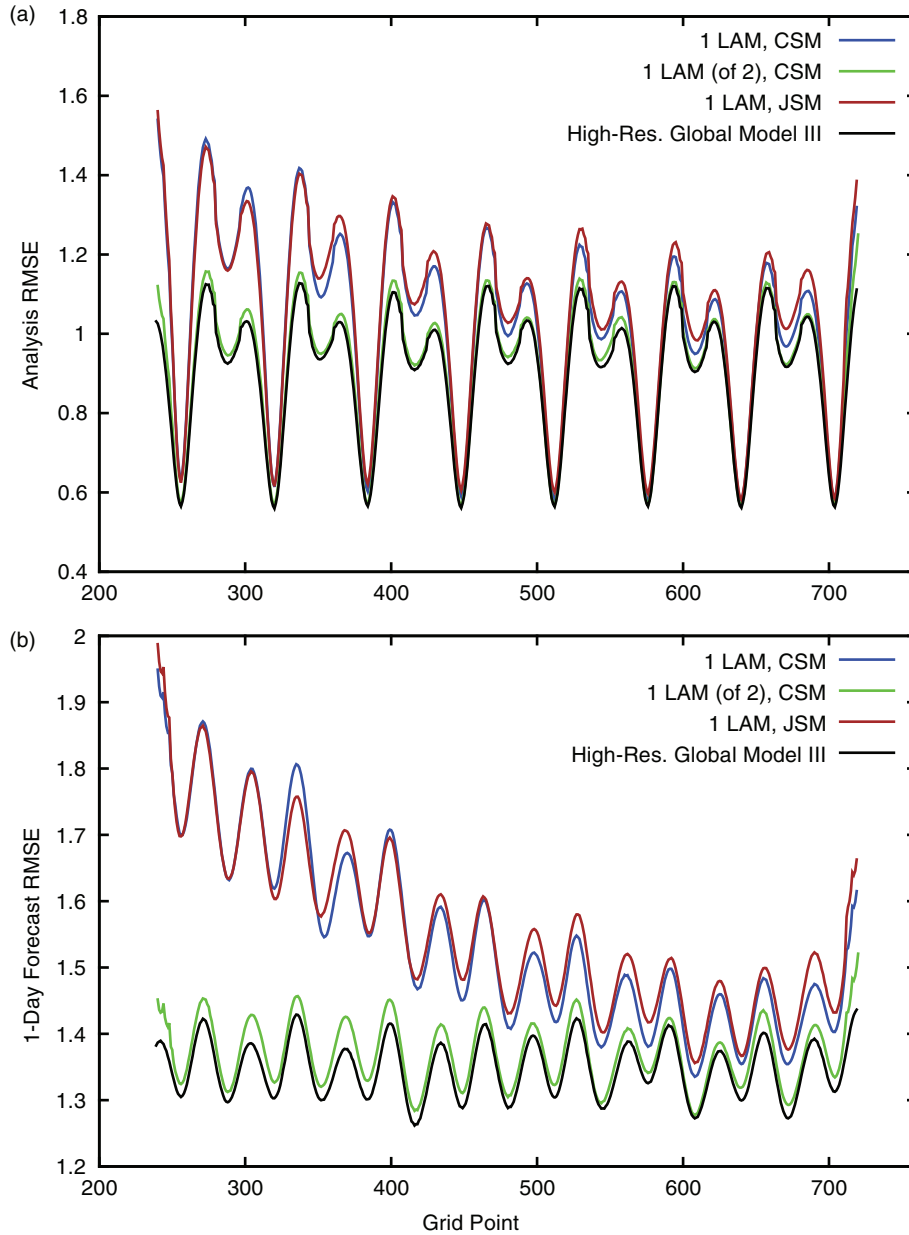


Fig. 8. (a) Ensemble mean analysis and (b) 1-d forecast accuracy gains with the addition of a second LAM. Statistics are gathered over 10^5 analysis cycles, after 10^3 cycles of spin-up time. In both the one and two LAM situations the LAM domain of interest runs from grid points [240,720]. The second LAM is added on the domain [720,240]. Curves denoted ‘CSM’ are calculated using the composite state method, and those denoted ‘JSM’ are found when using the joint-state method of Yoon et al. (2012). Single deterministic forecasts are initialised with the LAM analysis ensemble mean. The results from the perfect model ensemble are included over the LAM domain of interest as a benchmark for comparison.

An additional LAM also allows forecasts made by the lower resolution global model to improve, as shown in Fig. 9. Here we compare the 1-d forecast accuracy of forecasts produced by the global model, initialised with a composite state analysis mean calculated using 1 or 2 LAMs (blue and gold curves, respectively). The addition of an LAM helps to more accurately initialise global model forecasts across the

entire simulation domain, allowing forecast accuracies to approach those made by a higher resolution perfect model over the whole ‘globe’, rather than only at certain locations inside the LAM domain, which may be far from its boundaries.

Figure 10 shows results for our final experiment, which compares state estimates produced for two scenarios.

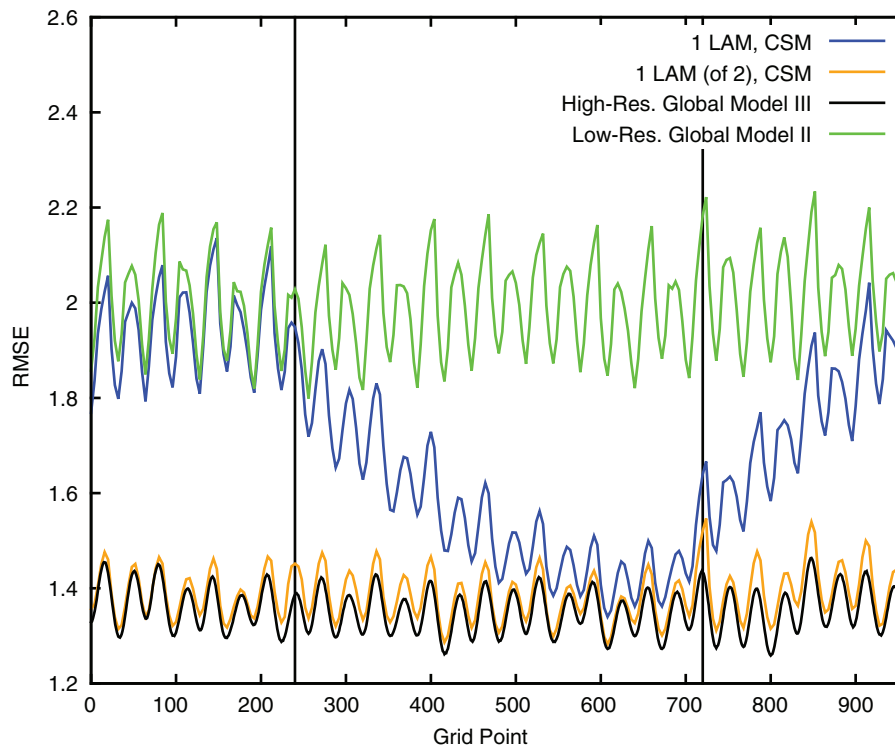


Fig. 9. Ensemble mean forecast accuracy of the global model, for the conditions described for Fig. 8. The addition of a second LAM dramatically lowers global model forecast accuracy over the entire global domain. Results from perfect and imperfect global model ensemble forecasts are included as a benchmark for comparison, and vertical black lines demarcate the LAM boundaries.

The first has two small overlapping LAMs, covering grid point intervals [240,500] and [460,720] (red curve), and the second has a single larger LAM defined over the union of these domains, the interval [240,720] (blue curve). Data assimilation is performed in both of these situations using the composite state method. Overall, the accuracy of the state estimates produced in both scenarios, when averaged over time, is almost the same. This is a somewhat surprising result, as the two LAMs are driven by the imperfect global dynamics. At regions of LAM overlap and near LAM boundaries, the composite state method eliminates much of the boundary condition errors that would otherwise be present at these locations (for example near the LAM boundaries in Fig. 1). Overall, we conclude from our experiments that by considering all relevant LAM state information during the analysis, the composite state method helps to decrease, and in some cases virtually eliminate, boundary condition errors, and that this improved analysis state translates to better forecast performance.

6. Summary and conclusions

While the advantages of using LAMs for short-term (about 48 hours and shorter) weather forecasting have been known for some time, data assimilation has been traditionally

performed solely on either the LAM or the global model that provides lateral boundary conditions. However, recent results indicate that this may not be the most optimal course of action (Guidard and Fischer, 2008; Dahlgren and Gustafsson, 2012; Yoon et al., 2012). Rather, forecast accuracy can increase by allowing the global model to influence LAMs, and vice versa, in the analysis. Motivated by these findings, and the fact that some large weather centres run LAMs for multiple regions, we present an ensemble data assimilation scheme that allows a global and several limited-area models to influence one another during the analysis procedure.

Using numerical experiments conducted under idealised test conditions, we show that our method has the potential to improve forecast accuracy. When applied to multiple LAMs, the first step of our analysis algorithm creates an ensemble of high-resolution ‘composite’ states from the ensembles of global and all limited-area model states. We then perform an ensemble analysis procedure (in the case considered in this paper, the LETKF), to arrive at a composite state analysis ensemble. Next, we use the composite state analysis ensemble to construct global and limited-area model ensembles to be employed as initial conditions for forecasts that provide the background ensemble for the next analysis cycle. We note that, through the analysis procedure

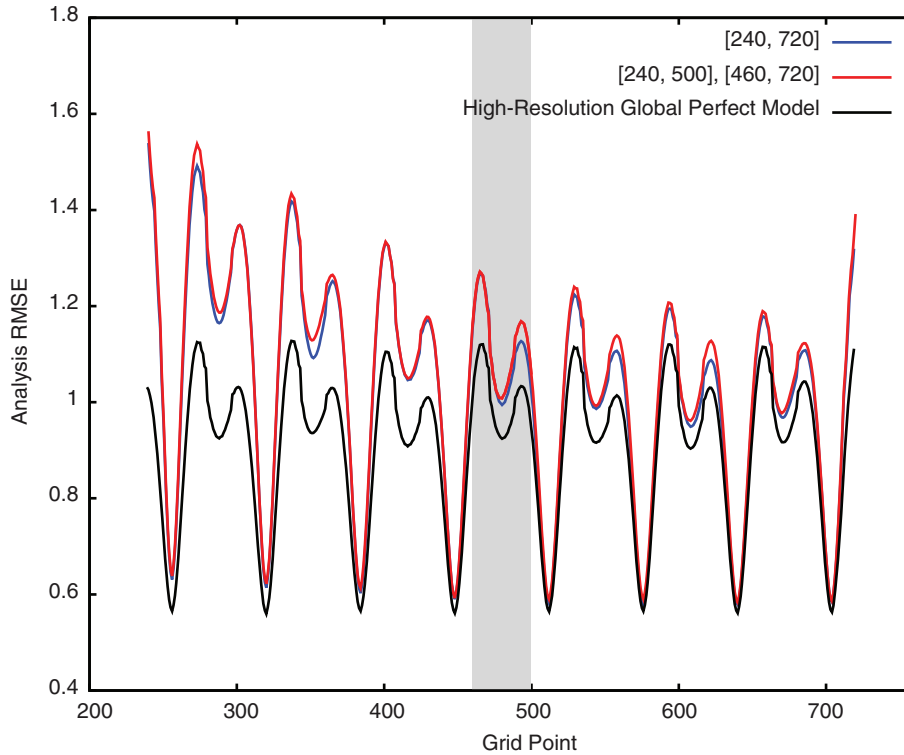


Fig. 10. RMS analysis errors of the ensemble mean calculated using the composite state method for two model scenarios. The first scenario has a single LAM defined over the grid point interval [240,720] (blue curve), and the second has two LAMs defined over the intervals [240,500] and [460,720] (red curve). The analysis error of the ensemble mean of a global high-resolution perfect model LETKF (black curve) is included as a benchmark for comparison. Statistics are gathered over 10^5 analysis cycles, discarding the first 10^3 cycles. The shaded area indicates the domain where both LAMs are defined.

and the boundary conditions supplied by the global model ensemble, the composite state method effectively allows observational information from outside of a LAM domain to influence the update of the estimate of the state at grid points inside of its domain. Additionally, we note that this general composite state method does not depend on the particulars of the LETKF, allowing for flexibility in the choice of the analysis algorithm.

For experiments with LAMs that cover the entire global domain, the composite state method is shown to be capable of producing forecasts with accuracies that are almost as good as those produced in the ideal case of assimilations using a global perfect model. Additionally, in situations where the collection of LAMs cover a subsection of the global domain, we show that there is a clear benefit to allowing the LAM states to influence one another during the data assimilation process. The experimental results shown here suggest that, in real weather forecasting situations, high-resolution state estimates provided by many (potentially overlapping) LAMs can be used to greatly improve global atmospheric state analysis estimates. Even including information about the state from LAMs with non-overlapping domains in the composite state analysis has

a positive effect on global model state estimates outside of a LAM domain (see blue and orange versus green curves in Fig. 9). The composite state method analysis technique could be a useful tool for organisations like the United States Navy, which currently runs the COAMPS[®] LAM for many, often overlapping, domains, as it presents a straightforward method for utilising a collection of disparate state estimates in the analysis. The composite state method would allow for short-term high-resolution forecasts produced by these regional models to improve the global analyses, in turn allowing for further improvements to the LAM forecasts through improved boundary conditions.

Our presentation of the composite state method has utilised simple one-dimensional models, each defined on subsets of the same grid. Atmospheric models are defined over three spatial dimensions, and different LAMs will not, in general, share common grid points, even if their domains happen to cover common geographic areas. Thus, in further tests of the composite state method on more realistic systems, it will be necessary to choose appropriate $p_i(n)$ functions, and to specify the interpolations that the operator O_n in eq. (2) is to perform. These choices will be dependent on the number and relative layout of LAM domains, and can be

empirically adjusted to yield the most desirable results. Our choice of p_i (n) used no information from the low-resolution forecast wherever a high-resolution forecast was available. It remains a possible subject for further study whether it is advantageous to retain some of this information in other scenarios. Overall, we are encouraged by our results from the one-dimensional models to speculate that the composite state method might offer a potential means of obtaining forecast improvements in real weather forecasting situations.

7. Acknowledgements

This work was supported by ONR grant N000141210785. COAMPS® is a registered trademark of the Naval Research Laboratory.

References

- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.
- Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.
- Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.
- Bonavita, M., Torrisi, L. and Marcucci, F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q. J. Roy. Meteorol. Soc.* **136**, 132–145.
- Bowler, N. E. and Mylne, K. R. 2009. Ensemble transform Kalman filter perturbations for a regional ensemble prediction system. *Q. J. Roy. Meteorol. Soc.* **135**, 757–766.
- Burgers, G., Jan van Leeuwen, P. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.
- Dahlgren, P. and Gustafsson, N. 2012. Assimilating host model information into a limited area model. *Tellus* **64A**, 18536. DOI: 10.3402/tellusa.v64i0.15836.
- Davies, H. C. 1983. Limitations of some common lateral boundary schemes used in regional NWP models. *Mon. Wea. Rev.* **111**, 1002–1012.
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, C5, 10143–10162.
- Guidard, V. and Fischer, C. 2008. Introducing the coupling information in a limited-area variational assimilation. *Q. J. Roy. Meteorol. Soc.* **134**, 723–735.
- Harris, L. M. and Durran, D. R. 2010. An idealized comparison of one-way and two-way grid nesting. *Mon. Wea. Rev.* **138**, 2174–2187.
- Holt, C. R., Szunyogh, I. and Gyarmati, G. 2013. Can a moderate-resolution limited-area data assimilation system add value to the global analysis of tropical cyclones? *Mon. Wea. Rev.* **141**, 1866–1883.
- Houtekamer, P. and Mitchell, H. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.
- Hunt, B. R., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* **230**, 112–126.
- Klocke, D. and Rodwell, M. 2014. A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Q. J. Roy. Meteorol. Soc.* **140**, 517–524.
- Lange, H. and Craig, G. C. 2014. The impact of data assimilation length scales on analysis and prediction of convective storms. *Mon. Wea. Rev.* **142**, 3781–3808.
- Lorenz, E. N. 2005. Designing chaotic models. *J. Atmos. Sci.* **62**, 1574–1587.
- Merkova, D., Szunyogh, I. and Ott, E. 2011. Strategies for coupling global and limited-area ensemble Kalman filter assimilation. *Nonlin. Proc. Geophys.* **18**, 415–430.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.
- Pielke, R. A. Sr. 2013. *Mesoscale Meteorological Modeling*. Vol. 98. Academic Press, Waltham, MA, USA.
- Reich, H., Rhodin, A. and Schraff, C. 2011. LETKF for the non-hydrostatic regional model COSMO-DE. *COSMO Newsletter* **11**, 27–31.
- Torn, R., Hakim, G. and Snyder, C. 2006. Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.* **134**, 2490–2502.
- Wang, X., Bishop, C. H. and Julier, S. J. 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.* **132**, 1590–1605.
- Warner, T. T., Peterson, R. A. and Treadon, R. E. 1997. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Met. Soc.* **78**, 2599–2617.
- Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.
- Yang, S.-C., Kalnay, E., Hunt, B. and Bowler, N. E. 2009. Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Q. J. Roy. Meteorol. Soc.* **135**, 251–262.
- Yoon, Y.-N., Hunt, B. R., Ott, E. and Szunyogh, I. 2012. Simultaneous global and limited-area ensemble data assimilation using joint states. *Tellus* **64A**, 18407. DOI: 10.3402/tellusa.v64i0.18407.
- Yoon, Y.-N., Ott, E. and Szunyogh, I. 2010. On the propagation of information and the use of localization in ensemble Kalman filtering. *J. Atmos. Sci.* **67**, 3823–3834.