Estimating forecast model bias in coupled global and limited-area models

By MATTHEW KRETSCHMER[1]*, BRIAN R. HUNT[2] and EDWARD OTT[1,3], [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*

ABSTRACT

Idealised perfect model experiments suggest that performing data assimilation on a 'composite' state vector, constructed from global and limited-area model states, can be beneficial to both model states. Here, an illustrative scheme is implemented to account for systematic forecast errors attributed to the imperfect model dynamics. Results from numerical experiments suggest that even simple bias correction schemes can correct forecast errors in composite states.

Keywords: ensemble Kalman Filter, limited-area models, composite state, bias correction

1. Introduction

Limited-area models are commonly used to generate regional high-resolution atmospheric forecasts. Correcting systematic forecast model errors, known as *forecast model bias*, can be complicated in such systems by the limited-area model's need for lateral boundary conditions. Lateral boundary conditions are typically provided by a concurrently running, nested limited-area or global model, each of which frequently exhibits model errors of its own. In operational forecasting, data assimilation is used to correct global and limited-area forecasts towards recent observations, but this is typically done separately for each model. Performing data assimilation simultaneously on both global and limited-area model states has recently been shown to be beneficial to both global and limited-area model states (Yoon et al., 2012; Kretschmer et al., 2015). However, simultaneous assimilation introduces additional coupling between model states, further allowing errors in each model to affect the other.

The composite state framework forms a state estimate from a combination of limited-area and global model forecasts (Kretschmer et al., 2015). The purpose of this paper is to illustrate the possibility of applying bias correction to the composite state method, and further, to illustrate that doing this may have substantial benefit. Correcting composite state forecast model bias accounts for the overall effect of model biases present in coupled global and limited-area forecast models. For this purpose, we will use a bias correction scheme initially presented in Baek et al. (2006).

Data assimilation techniques have been successfully applied to estimate the effect of, and to correct for, forecast model biases using variational and ensemble techniques (Dee and Da Silva, 1998; Carton et al., 2000; Dee, 2005; Keppenne et al., 2005; Li et al., 2009). Often, this is accomplished through state-vector augmentation techniques (Jazwinski, 1970; Cohn, 1997; Anderson and Anderson, 1999). The technique presented here aims to estimate the cumulative effect of model error on short-term model forecasts. Note that this bias correction method only modifies the analysis procedure, and not the forecast model equations. While many studies explore and account for the effects of forecast model bias on global models, many fewer have investigated how to account and correct for forecast model bias when limited-area models are involved [e.g. see the review by Meng and Zhang (2011)].

2. Background

We test our bias correction scheme in numerical experiments with biased forecast models, similar to Baek et al. (2006).

---

*Corresponding author.
email: mkretsch@umd.edu

1

For this discussion, and in our experiments, the 'truth' model dynamics are denoted as

$$\frac{d\mathbf{x}^T}{dt} = \mathbf{M}(\mathbf{x}^T), \tag{1}$$

and the vector $\mathbf{x}^T$ represents the truth model state vector.[1] We model forecast bias as an additive model error tendency; details of the specific biased forecast models used in our experiments are presented below.

2.1. Forecast model bias correction

The present work adapts the composite state method to correct for forecast model errors. Baek et al. (2006) presented a series of 'bias models' to account for forecast bias resulting from imperfect forecast model dynamics in ensemble forecasting systems. Data assimilation is performed on a chosen bias model to adaptively estimate both the model state and forecast bias. This is accomplished by augmenting the state vector $\mathbf{x}$ with a vector of 'bias corrections'.

Here, we consider Bias Model I proposed by Baek et al. (2006). In particular, this bias model assumes that, after spin-up, forecasts are initialised at time t_{m-1} to the approximate truth state at t_{m-1}, $\mathbf{x}^T_{m-1}$. Forecast errors then arise because of deviations between the truth and forecast model dynamics. Using data assimilation, the unknown vector $\mathbf{b}$ is estimated so that the truth at time t_m is estimated by Bias Model I

$$\mathbf{x}^T_m \approx \mathbf{F}(\mathbf{x}^T_{m-1}) + \mathbf{b}, \tag{2}$$

where $\mathbf{F}$ is obtained by integrating the (biased) forecast model from time t_{m-1} to time t_m. We note that the operator $\mathbf{F}$ may operate on either a model state estimate or the true state. The analysis is performed on the augmented state vector $\mathbf{x}$, consisting of the corrected forecasted state vector, $\mathbf{x}^b_m = \mathbf{F}(\mathbf{x}^a_{m-1}) + \mathbf{b}^b_m$ and the current bias correction estimate $\mathbf{b}^b_m$. Both model state and bias correction estimates can then be updated using observations and the covariances between model variables and bias correction estimates to yield updated analysis estimates $\mathbf{x}^a_m$ and $\mathbf{b}^a_m$, which may then be forecast to the next analysis time, t_{m+1}. The vector of bias corrections is evolved to time t_{m+1} via $\mathbf{b}^b_{m+1} = \mathrm{G}_b(\mathbf{b}^a_m)$. The exact form of the bias correction time evolution operator G_b depends upon the properties of the detected forecast errors and may be empirically tuned as necessary.

2.2. Composite state forecasting

We account for forecast model bias in a coupled system of global and limited-area models. Data assimilation is performed on a background composite state vector, assumed to provide an optimal state estimate (Kretschmer et al., 2015). This composite state is formed by combining information from the global and limited-area model forecasts (which include coupling of global model provided boundary conditions), and is essentially created from the highest spatial resolution forecast information available at a given geographical location. Upon completion of the analysis procedure, the analysis composite state estimate $\mathbf{x}^a_{m-1}$ is evolved to time t_m using

$$\mathbf{x}^b_m = \mathbf{L}(\mathbf{x}^a_{m-1}) + \mathbf{b}^b_m. \tag{3}$$

The operator $\mathbf{L}$ initialises the global and limited-area model states to $\mathbf{x}^a_{m-1}$, evolves these states from time t_{m-1} to t_m using the (biased) global and limited-area forecast models and from these forecasts forms the composite state $\mathbf{x}^b_m$. For additional details regarding the construction of the composite state vector and the form of $\mathbf{L}$ used below, see Kretschmer et al. (2015).

3. Composite state bias correction

We consider coupled data assimilation performed between limited-area and global model states. When forecast models and data assimilation are coupled, errors in either of the forecast models can affect both limited-area and global model forecasts. Instead of accounting for model error separately in each of these forecast models, we consider errors in the composite state forecast, which represent cumulative effects of limited-area and global forecast model biases.

We approximate the effect of errors in the composite state forecast model, $\mathbf{L}$ in eq. (3), with an additive correction, using the composite state vector when applying eq. (2). Here $\mathbf{x}^T_m$ represents the truth at time m, and both $\mathbf{b}^b_m$ and $\mathbf{x}^T_m$ have the same spatial resolution as the composite state. As we shall see, correcting the forecast model biases of the composite state system allows the composite state, the best estimate of the true system state, to be dramatically more accurate in the presence of forecast model error than it would otherwise be.

4. Experimental details

4.1. The Lorenz models

Our numerical experiments use the one-dimensional chaotic models of Lorenz (2005). The dynamics of Lorenz's Model II is given at grid point n by

$$\frac{dZ_n}{dt} = [Z, Z]_{K,n} - Z_n + F. \tag{4}$$

The bracket term in eq. (4) represents spatially smoothed nonlinear advection, and F is a constant forcing parameter. Specifically, the bracket term of eq. (4) depends upon a

---

[1]In general, the dynamics represented by the operator M may depend explicitly on time.

smoothing parameter K and can be calculated for two quantities X and Y at each grid point n using

$$[X, Y]_{K,n} = \frac{1}{K^2} \sum_{j=-J}^{J}{}' \sum_{l=-J}^{J}{}' (X_{n-K+j-l} Y_{n+K+j} - X_{n-2K-l} Y_{n-K-j}), \qquad (5)$$

In eq. (5), $J = K/2$ if K is even and $J = (K-1)/2$ if K is odd (Lorenz, 2005). Lorenz's Model III models the state variable Z_n as the sum of long and short spatial scale components, with $Z_n = X_n + Y_n$. The long scale component X_n is defined as an average of Z_n over nearby grid points, and the short scale component Y_n is defined as the residual $Y_n = Z_n - X_n$. The dynamics of Model III are given by

$$\frac{dZ_n}{dt} = [X, X]_{K,n} + d^2 [Y, Y]_{1,n} + c[Y, X]_{1,n} - X_n - dY_n + F. \quad (6)$$

In the above model equations, d, c and K are parameters, and t represents time. See Lorenz (2005) for more details of these models.

For our numerical experiments, Model II describes the global model dynamics, and Model III describes both limited-area and truth model dynamics. The global and limited-area models are defined on different subsets of the $N = 960$ grid point lattice on which the truth model state is defined. The grid points of the truth lattice are indexed from 0, and all model state values are referenced by their index on the truth grid. The global model state is defined on every fourth point of the $N = 960$ truth grid, and the limited-area model is defined on a subset of the truth model grid, over the grid point interval [240, 720].

The 'true' state is generated from a long free model run of Model III with parameters $d_T = 10$, $c_T = 0.6$, $F_T = 15$ and $K_T = 32$. In addition to the inherent bias of the global model due to its lower spatial resolution, we also introduce a bias by shifting the value of F used in the limited-area and global forecast models from the value used in the truth model, F_T. The global model uses $K_g = 8$ to match the length scale of $K = 32$ used in Model III. Below, we use $\mathbf{x}_m$ to represent the values of Z at time m at the composite state
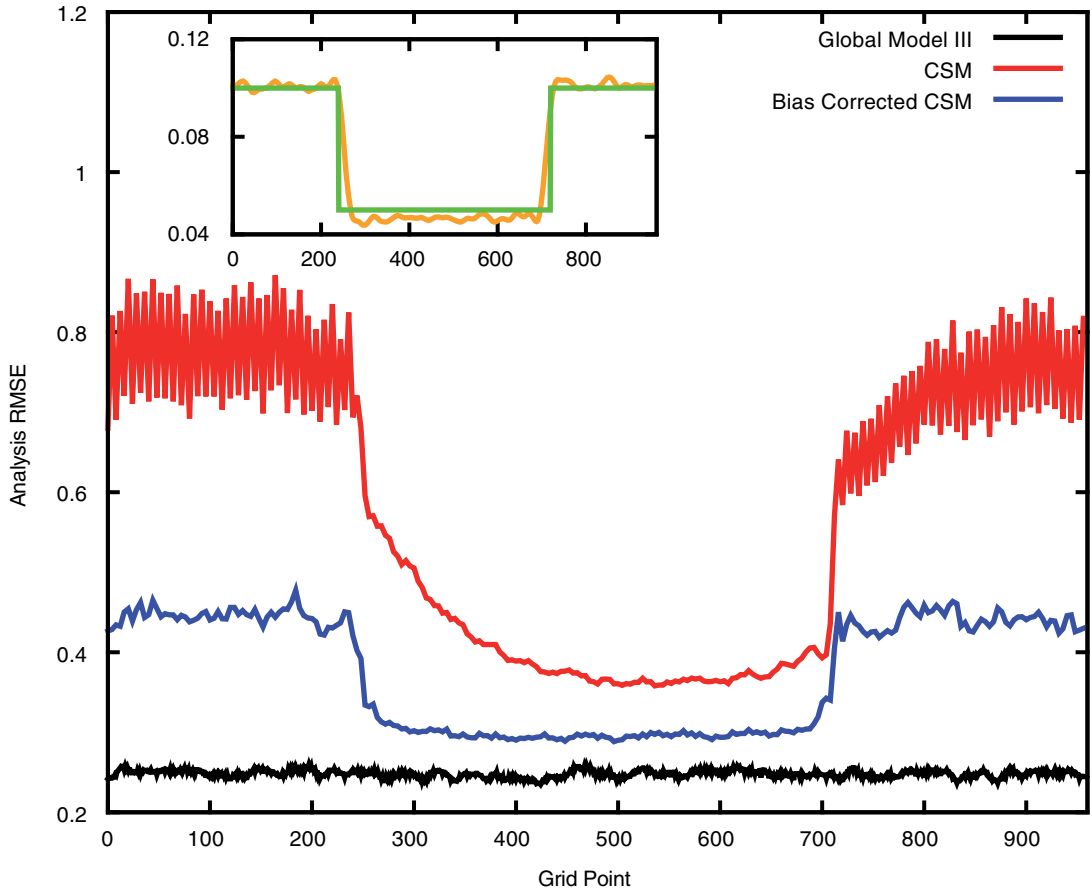

Fig. 1. RMSE of the analysis ensemble mean [eq. (8)] when using the composite state method (CSM). Bias correction (blue curve) significantly increases the analysis accuracy compared with the analysis without bias correction (red curve), and approaches perfect (unbiased) global forecast model results (black curve). The inset figure plots the spatial dependence of the time-averaged estimated bias correction $\mathbf{b}$ (gold curve) and the estimate provided by eq. (9) (green curve).

grid points, consisting of every grid point in [240, 720] and every fourth grid point elsewhere.

4.2. Data assimilation

In the numerical experiments, data assimilation is performed on a 32-member composite state ensemble using the Local Ensemble Transform Kalman Filter (Ott et al., 2004; Hunt et al., 2007), with a local analysis patch size of 81 grid points. The initial composite state ensemble is generated from samples of a free run of the global Model II. To prevent filter divergence, constant multiplicative covariance inflation is utilised (Anderson and Anderson, 1999). We achieve optimal results when composite state vector and bias correction parameter ensembles are inflated independently, by different amounts.

Observations are created by adding Gaussian white noise with mean 0 and standard deviation 1 to the true state at observation locations, and are directly compared to the value of the composite state at the observation location. The numerical experiments utilise a homogeneous observa-

tion network, with observations generated and assimilated every eight grid points.

4.3. Bias correction

We find that evolving the bias corrections in time using weak, numerical spatial diffusion allows for larger time steps and smaller ensemble sizes, in addition to speeding convergence, as previously reported in Baek et al. (2006). In our experiments, **b** is evolved in time via

$$\mathbf{b}_{m+1}^b(n) = (1 - 2D_b(n))\mathbf{b}_m^a(n) + D_b(n)\mathbf{b}_m^a(n-1) \\ + D_b(n)\mathbf{b}_m^a(n+1). \tag{7}$$

The parenthetical notation (n) in eq. (7) denotes the value of the given field at grid point n. We empirically tune the diffusion coefficient $D_b(n)$ over a range of magnitudes in order to minimise the root mean square (RMS) analysis error for each forecast model bias. We choose the diffusion coefficient $D_b(n)$ to be spatially constant where the composite state vector has constant spatial resolution.
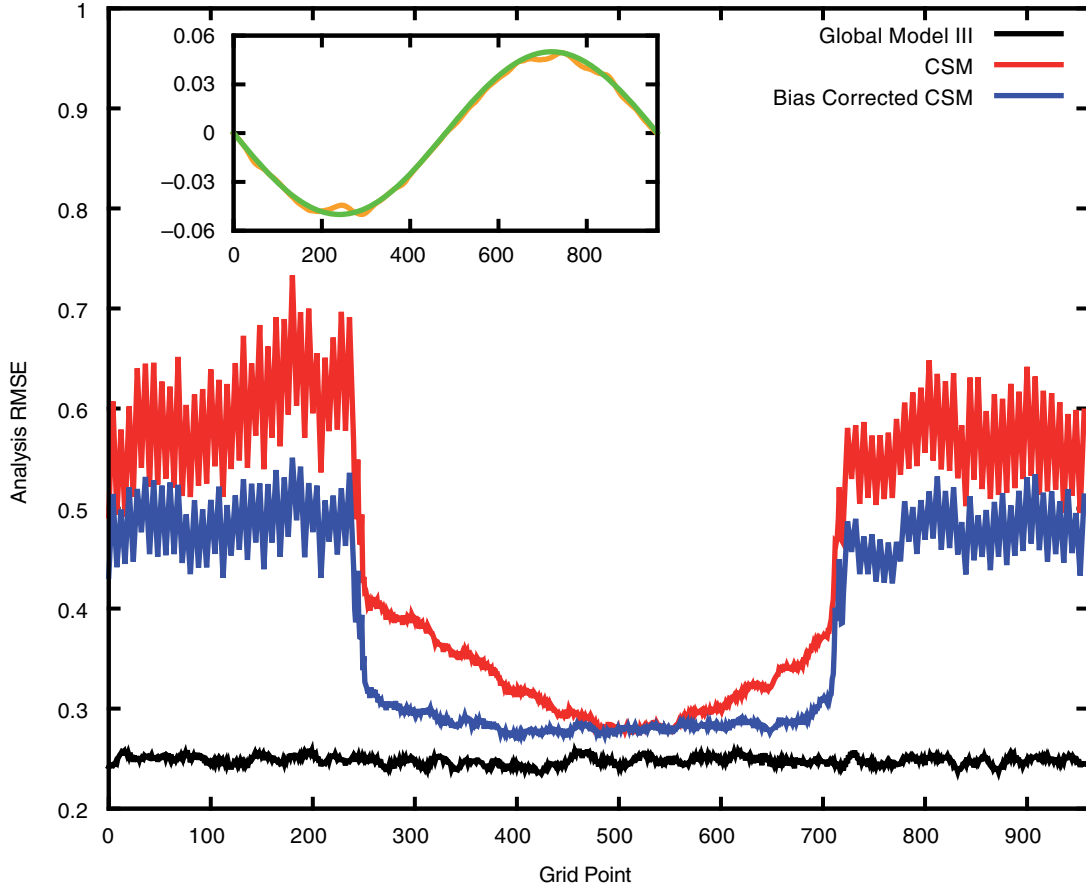

Fig. 2.   Same as Fig. 1, but with spatially dependent bias ΔF. Bias correction leads to decreased analysis RMSE compared to the composite state analysis without bias correction (blue versus red curves, respectively). The inset figure shows averaged bias correction **b** (gold curve) versus the value predicted by eq. (9) (green curve).

In all experiments, the components of the forecast model bias are constant in time. When the global and limited-area forecast models are biased differently, their model biases are denoted with the subscripts g and r, respectively.

4.4. Error metric

We use the root mean square error (RMSE) of the analysis ensemble mean as a metric of analysis accuracy, and the root-mean square error of the forecast ensemble mean to measure forecast accuracy. Ensemble forecasts are initialised from the analysis ensemble. We define $\epsilon(m, n, f) = \overline{\mathbf{x}}_m^f(n) - \mathbf{x}_m^T(n)$ as the difference at location n and time m between the f-hour lead time forecast ensemble mean and the truth, respectively. The RMSE averaged over time is

$$RMSE(n, f) = \left\{ \sum_{m=1}^{c} (\epsilon(m, n, f))^2 / c \right\}^{1/2}. \tag{8}$$

The RMSE of the analysis ensemble corresponds to $f = 0$ in eq. (8). Unless noted, results shown use $c = 20000$ in eq. (8).

5. Results and discussion

Our first numerical experiment uses Lorenz Model III [eq. (6)] for the limited-area model dynamics with parameter values $d_r = 10$, $c_r = 0.6$, $K_r = 32$ and forcing $F_r = 14$, so that the limited-area model dynamics may be denoted as the truth dynamics with an additive bias given by $\Delta F_r = F_r - F_T = -1$. We use Lorenz Model II [eq. (4)] with $K_g = 8$ and forcing $F_g = 13$ to govern the global model dynamics. The values of d_r, c_r, K_r and K_g are constant across all experiments described here. The RMSE of the analysis composite state ensemble mean, calculated for constant forecast model bias, is shown in Fig. 1. It is seen that the composite state method is able to achieve significantly improved results when applying bias correction (blue curve) compared with the non-bias corrected case (red curve).

The bias-corrected composite state analysis corrects for both the spatially independent global model bias, $\Delta F_g = F_g - F_T = -2$, and the lower resolution and imperfect global model dynamics (Model II). We note that, without bias correction, the limited-area model has substantial
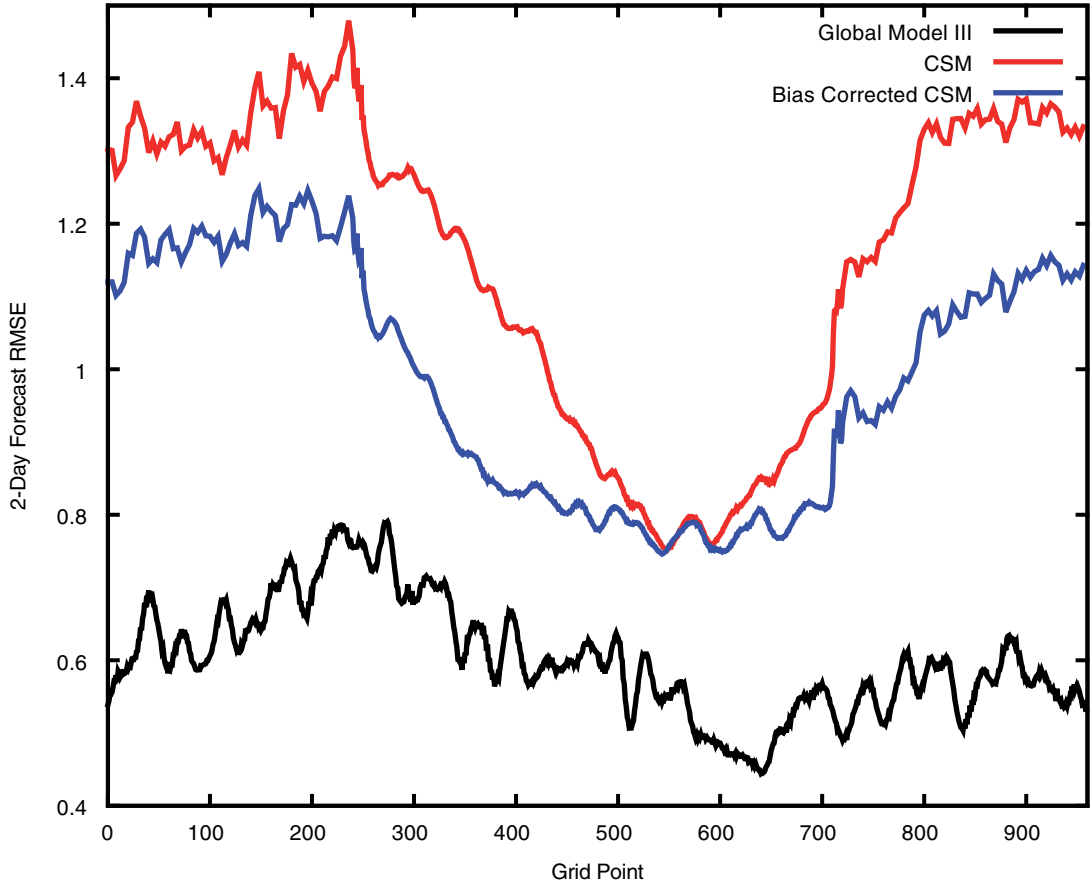

Fig. 3. RMSE of 2-d forecast ensemble mean, for spatially dependent bias as in Fig. 2. Blue and red curves compare forecasts with and without bias correction, respectively. The black curve shows ensemble forecast RMSE when forecasting with the truth model.

error near the left boundary of its domain, and this feature is not seen with bias correction. We interpret this as evidence that the bias-corrected composite state improves lateral boundary conditions to the limited-area model, as the Lorenz models exhibit rightward information propagation (Yoon et al., 2010).

When modelling forecast model bias with additive model error tendencies, as is the focus of this manuscript, we expect the bias corrections to converge to

$$\mathbf{b} \approx -\Delta F \Delta t, \qquad (9)$$

where Δt denotes the assimilation window (Baek et al., 2006), which for our experiments has the value $\Delta t = 0.05$. For the constant model biases in this experiment, eq. (9) predicts $\mathbf{b}$ should converge to $\mathbf{b} \approx 0.1$, where the global model is defined and $\mathbf{b} \approx 0.05$, where the limited-area model is defined. The bias correction parameters estimated in this experiment are shown in the inset of Fig. 1 (gold curve), and closely approximate the predicted asymptotic values of $\mathbf{b}$ shown in green.

In practice, it is likely that model errors will have some spatial dependence, and to investigate how well such biases may be corrected for in composite state forecasts, we allow the forecast model biases to vary according to $\Delta F_g(n) = \Delta F_r(n) = \sin(2\pi \frac{n}{960})$. Figure 2 shows the RMS analysis error of the composite state when there is spatially varying forecast model bias. As in Fig. 1, the bias controlled composite state analysis (blue curve) again outperforms the uncorrected composite state analysis (red curve). Bias correction accounts very well for the imposed model error in ΔF as well as the improper global forecast model dynamics, as evidenced by the relatively flat behaviour of the analysis RMSE curve.

The increased accuracy of the bias-corrected analysis implies that the effect of the model bias is being accurately estimated. According to eq. (9), the bias corrections are expected to converge to $\mathbf{b} = -0.05 \sin(2\pi \frac{n}{960})$, and this estimate is plotted in the inset of Fig. 2 as a green curve, along with the time-averaged value of $\mathbf{b}$ (gold curve). Their close agreement illustrates how the bias correction scheme is effective when forecast model bias is spatially varying.

Bias correction can potentially improve forecast results as well. We consider forecast lead times that are multiples of the assimilation window, Δt. Every Δt hours after initialisation at time t_m, each forecast ensemble member $\mathbf{x}^F$ is adjusted according to $\mathbf{x}^F \rightarrow \mathbf{x}^F + \mathbf{b}_m^a$. Figure 3 shows RMSE of 2-d forecasts resulting from integrating an ensemble of global model states, initialised from the composite state analysis ensemble, while applying this methodology. Correcting the forecast model bias, even in

this rudimentary fashion, clearly improves forecast accuracy, even out as far as 2-d lead times.

Current atmospheric forecast models contain myriad possible sources of error that can contaminate forecasts. When forecasting with coupled limited-area and global forecast models, these errors can affect the output forecasts in complicated and non-trivial ways, especially when performing coupled data assimilation on limited-area and global model states. Using a simple illustrative setup, we present here evidence that an appropriate method can account for cumulative forecast model errors in a coupled forecasting system, and that our results suggest that the composite state method with bias correction may be useful for producing forecasts with imperfect model dynamics.

6. Acknowledgements

References

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

Baek, S.-J., Hunt, B. R., Kalnay, E., Ott, E. and Szunyogh, I. 2006. Local ensemble Kalman filtering in the presence of model bias. *Tellus A.* **58**, 293–306.

Carton, J. A., Chepurin, G., Cao, X. and Giese, B. 2000. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: methodology. *J. Phys. Ocean.* **30**, 294–309.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **75**, 147–178.

Dee, D. P. 2005. Bias and data assimilation. *Q. J. Roy. Meteorol. Soc.* **131**, 3323–3343.

Dee, D. P. and Da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Q. J. Roy. Meteorol. Soc.* **124**, 269–295.

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.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory.* Academic Press, San Diego, CA.

Keppenne, C., Rienecker, M., Kurkowski, N. and Adamec, D. 2005. Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Non. Proc. Geophys.* **12**, 491–503.

Kretschmer, M., Hunt, B., Ott, E., Bishop, C. H., Rainwater, S, and co-authors. 2015. A composite state method for ensemble data assimilation with multiple limited-area models. *Tellus A.* **67**, 26495, doi: http://dx.doi.org/10.3402/tellusa.v67.26495

Li, H., Kalnay, E., Miyoshi, T. and Danforth, C. M. 2009. Accounting for model errors in ensemble data assimilation. *Mon. Weather Rev.* **137**, 3407–3419.

Lorenz, E. N. 2005. Designing chaotic models. *J. Atmos. Sci.* **62**, 1574–1587.

Meng, Z. and Zhang, F. 2011. Limited-area ensemble-based data assimilation. *Mon. Weather Rev.* **139**, 2025–2045.

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 A*. **56**, 415–428.

Yoon, Y.-N., Hunt, B. R., Ott, E. and Szunyogh, I. 2012. Simultaneous global and limited-area ensemble data assimilation using joint states. *Tellus A*. **64**, 18407. doi: http://dx.doi.org/10.3402/tellusa.v67.18407

Yoon, Y.-N., Ott, E. and Szunyogh, I. 2010. On the propagation of information and the use of localisation in ensemble Kalman filtering. *J. Atmos. Sci.* **67**, 3823–3834.