

New method for estimating the impact of heterogeneous forcing on atmospheric circulations

Tomislava Vukicevic,^{1,2} Roger A. Pielke Sr.,^{1,2} and Adriana Beltrán-Przekurat^{1,2}

Received 14 May 2008; revised 21 November 2008; accepted 31 December 2008; published 3 April 2009.

[1] In this study, a new method for estimating the impact of heterogeneous forcing on atmospheric circulations is discussed. This new method is similar to the commonly used model-based sensitivity studies in that the impact of forcing is diagnosed by a suitable measure of differences between atmospheric states with and without forcing, but it differs in the way the atmospheric states are evaluated: by combining standard atmospheric data analysis, observationally based estimates of the forcing, atmospheric observations, and general circulation model (GCM) ensemble simulations. A new numerical technique, derived from the ensemble Kalman filter data assimilation approach, is used for objective estimation of the atmospheric state not affected by the forcing. Using a tutorial example, numerical experiments were conducted varying an asymmetric thermal forcing as a proxy for the heterogeneous forcing. Results show that the method is capable of producing skilled estimates of the impact of the forcing. Strategies for application of the method with real-world data and GCMs are discussed. This new method is expected to produce more realistic estimates of the forcing impact than the standard model sensitivity approach because of the explicit use of the observationally based estimates of atmospheric states and forcing.

Citation: Vukicevic, T., R. A. Pielke Sr., and A. Beltrán-Przekurat (2009), New method for estimating the impact of heterogeneous forcing on atmospheric circulations, *J. Geophys. Res.*, 114, D07103, doi:10.1029/2008JD010418.

1. Introduction

[2] Understanding the impact of temporally and spatially heterogeneous forcing mechanisms on atmospheric circulations is of considerable interest because such forcing could potentially significantly contribute to the variability of climate on all time and spatial scales [National Research Council (NRC), 2005; Matsui and Pielke, 2006; Rial et al., 2004; Strack et al., 2008]. Notable examples of potentially significant mechanisms are radiative forcing by aerosols and clouds and variations in surface fluxes due to changes in land properties [NRC, 2005; Kabat et al., 2004].

[3] The response of atmospheric circulations to heterogeneous forcing mechanisms is difficult to assess from observations alone because the observations by definition include the effect of multiscale and multiprocess interactions and feedbacks which are difficult to decouple. Instead, the impact of forcing is often analyzed by model sensitivity studies where simulated atmospheric states with different parameterization and/or versions of the forcings are compared to a baseline simulated state without the forcings (for example, for aerosol radiative forcing effects on climate, see

Kim et al. [2007], Ramanathan et al. [2005], and Menon et al. [2002]).

[4] The intercomparisons that have been completed so far generally focus on global average radiative forcing as emphasized in the summary for policy makers of *Intergovernmental Panel on Climate Change* [2007]. Regional evaluations of the forcings have been completed but primarily in terms of maps of their effect on the spatial pattern of radiative forcing (vertical and horizontal) as shown, for example, in the work of *Climate Change Science Program* [2006]. While useful, these spatial assessments do not take the next step which is to evaluate the relative role of each of the diversity of climate forcings identified in NRC [2005] responsible for altering regional circulation patterns.

[5] The impact of forcing (I) can be represented by the following general relationship:

$$I \equiv \Gamma(g(x_{\text{true}}) - g(x_{\text{no forcing}})), \quad (1a)$$

where x_{true} and $x_{\text{no forcing}}$ are the true state of the atmosphere with and without the forcing respectively, $g(x)$ is a diagnostic of the state of the atmosphere and the function Γ is a suitable measure of distance of that diagnostic state. In a typical model-based impact assessment the true state of the atmosphere are approximated by the corresponding simulated states with and without forcing, x_{forcing}^* $x_{\text{no forcing}}^*$:

$$I \cong \Gamma(g(x_{\text{forcing}}^*) - g(x_{\text{no forcing}}^*)). \quad (1b)$$

¹Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado, USA.

²Also at Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, Colorado, USA.

For example, in the work of *Kim et al.* [2007], Γ is, the time average and $g(x)$ is large-scale circulation pattern or rainfall within a region in South Asia. When the model is erroneous, neither x_{true} nor $x_{\text{no forcing}}$ would be well approximated. Additional errors could result from boundary conditions if these are important for the particular model and the choice of the diagnostic $g(x)$. It is important to notice that contrary to what is typically assumed in model sensitivity studies, boundary conditions are not necessarily the same for x_{forcing}^* and $x_{\text{no forcing}}^*$ because the atmosphere with and without the forcing would be consistent with a different state at the time and location of these conditions.

[6] The model solution x_{forcing}^* could be improved by using observationally based estimates of heterogeneous forcing instead of the parameterization with free parameters, and by observationally based initial and lateral boundary conditions. The latter two could be specified, for example, from atmospheric reanalysis data. This approach has been used in studies on the impact of land use changes and aerosols (e.g., as summarized by *Kabat et al.* [2004] and *NRC* [2005]). Reanalyses have already been used very effectively to investigate the role of landscape changes and urbanization on long-term temperature trends [*Kalnay et al.*, 2006; *Cai and Kalnay*, 2005]. An implicit assumption in the studies with the observationally based forcing in the model is that the time scale at which the impact is measured (i.e., the time scale in $\Gamma[g(x)]$) is shorter than the time scale of significant feedbacks of the atmospheric circulation on the forcing. The model solution without forcing $x_{\text{no forcing}}^*$ is obviously not affected by the specification of the forcing, but the inconsistency between this and the observationally based boundary conditions may be even stronger.

[7] In this study, an alternative method for estimating the impact of forcing is proposed where the true state under the influence of forcing (x_{true}) and the state without the forcing ($x_{\text{no forcing}}$) are approximated by observationally based estimates instead of by a model alone. The purpose of such an approach is to reduce the influence of model errors and inconsistency between the state and initial and lateral boundary conditions. The new method, designated impact of forcing by prior state estimating (IFPRE), is based on the ensemble Kalman filter (EnKF) data assimilation approach, which is one of the commonly used approaches in atmospheric and other environmental data analysis [*Kalnay*, 2003; *Lewis et al.*, 2006; *Evensen*, 2006].

2. Description of the Impact of Forcing by Prior State Estimating (IFPRE) Method

[8] To obtain values of the forcing impact that correspond to the actual atmosphere, the states x_{true} and $x_{\text{no forcing}}$ in the expression (1a) should be approximated as close as possible to the true atmosphere with the equivalent forcing conditions. The true state of the atmosphere cannot be known, but a reasonably good approximation can be obtained by objective observationally based estimates. These estimates are available from standard atmospheric analysis or reanalysis data which are produced by operational weather and climate data centers such as European Center for Medium-Range Weather Forecast (ECMWF), National Center for Environmental Prediction (NCEO), and NASA. These anal-

yses represent the atmosphere which has undergone the influence of forcing. As such, the atmospheric reanalysis data should be used to approximate x_{true} in expression (1a). Next, we present a new method, the IFPRE method, to obtain observationally based estimates of the atmosphere without forcing $x_{\text{no forcing}}$ using the EnKF data assimilation approach.

[9] The atmospheric data assimilation by the EnKF approach combines data from observations and an ensemble of model simulations to objectively estimate either the state of atmosphere or the state together with a set of model parameters. The underlying mathematical theory of the EnKF methods and their utility in wide range of applications such as atmospheric, oceanic and other environmental daily data analysis [*Kalnay*, 2003; *Evensen*, 2006], climate modeling [*Annan et al.*, 2005] and climate sensitivity [*Hargreaves and Annan*, 2006] is summarized in Appendix A. Here we make use of the expression for the estimate of statistical mean state together with a set of model parameters, which is written

$$\begin{pmatrix} x_e \\ \alpha_e \end{pmatrix} = \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix} + \begin{bmatrix} K_x \\ K_{x\alpha} \end{bmatrix} \left(y - [H \ 0] \cdot \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix} \right), \quad (2)$$

where x_e and α_e are the state and parameter posterior estimates, respectively.

[10] The estimation by the expression (2) requires prior knowledge of the state and parameter statistical mean values, denoted by the vector

$$\begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix},$$

and the weighted differences between the observations y and projected prior knowledge of the state x_{prior} into the observation space. The weights are provided by the Kalman gain matrix

$$\begin{bmatrix} K_x \\ K_{x\alpha} \end{bmatrix}.$$

This matrix is derived from the standard estimation theory when it is assumed that the observations and prior have approximately Gaussian errors [*Tarantola*, 2005; *Evensen*, 2006]. The expression for the Kalman gain matrix is provided in Appendix A.

[11] In the EnKF formulation the Kalman gain matrix is evaluated from ensemble-based estimates of error covariances of the prior data and the observation error covariance (see Appendix A). The operator H in the expression (2) is a mapping from the model state into observation space. For example, H may be the spatial interpolation from the model grid point values of quantities in x_{prior} to the observation location if the same quantities are observed. The expression

$$[H \ 0] \cdot \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix}$$

in (2) indicates that only the modeled state is mapped into the observation space and not the model parameters.

[12] Assuming that the parameter vector α contains a quantitative representation of a heterogeneous forcing field, written

$$\alpha \equiv f(s, t),$$

where $f(s, t)$ is a function of space, denoted s , and time, denoted t , the EnKF data assimilation solution by the expressions (2) would produce estimate of the statistical mean of the forcing field $f(s, t)$ together with the corresponding state, given the atmospheric observations. This property implies that the EnKF approach provides an explicit relationship between the mean state prior to the influence of forcing and the forcing estimate. From the expression (2), this is written

$$f_e = f_{\text{prior}} + K_{xf}(y - Hx_{\text{prior}}), \quad (3a)$$

with

$$f_{\text{prior}} = 0, \quad (3b)$$

where f_e is the estimate of the heterogeneous forcing and K_{xf} is a component of the Kalman gain matrix which contains the ensemble-based estimate of the error cross-correlation between the forcing and state (see Appendix A). The optimal mean forcing estimate should be approximating the actual forcing. Assuming that the actual forcing is available from the observationally based estimates without the data assimilation, denoted f_{obs} , the relationship (3a) could be then used to estimate the associated prior mean state without that forcing, denoted x_{prior} in the expression (3a). This could be achieved by solving the following least square problem:

$$\{D(x_{\text{prior}})\}_{\min}, \quad D \equiv (K_{xf}(y - Hx_{\text{prior}}) - f_{\text{obs}})^T \cdot (K_{xf}(y - Hx_{\text{prior}}) - f_{\text{obs}}), \quad (4)$$

where $()^T$ denotes transpose of the vector. D is the sum of square differences between the estimated and observed values of the forcing. The state (x_{prior}) in the expression (4) is arbitrary and would depend on the problem of interest. Using hypothetical example of analysis of aerosol radiative forcing impacts, equivalent to studies by *Kim et al.* [2007], *Ramanathan et al.* [2005], and *Menon et al.* [2002], the atmospheric state to be estimated may include only a long-term time average of a regional circulation field. The observation data (y) may be also filtered to the equivalent temporal and spatial scales for consistency in information content.

[13] It is important to notice that the minimum of D , denoted x_{prior} , is unique because the relationship between x_{prior} and f_e by (3a) is linear, implying that D is exactly quadratic in x_{prior} . The minimum solution is the pseudo inverse of the relationship (3a). The direct inverse cannot be computed because the matrix $K_{xf}H$ is not square in general due to the difference in size of the forcing and state vectors. The state vector is typically significantly larger than the forcing vector.

[14] The minimization of D requires a first guess for the state x_{prior} , the cross-covariance matrix K_{xf} , observations y and the mapping function H . The first guess could be

provided by the model simulation or standard atmospheric reanalysis. Owing to the property that D is exactly quadratic the minimum solution would not depend in theory on the first guess. The estimates of K_{xf} could be readily obtained from the ensemble simulations using any state of the art general circulation model (GCM). The model ensemble for this purpose must include perturbations in the forcing which would match the spatial and temporal structure of the actual forcing as represented by the observationally based estimates. For example, observationally based estimates of aerosol radiative forcing with spatial structure are shown by *Matsui and Pielke* [2006]. The observations y and matrix H could be obtained from the atmospheric reanalysis system. The observations y could be also approximated by the reanalysis data because these data are created to be the best available estimates of the actual state. In this case the observation vector would include the reanalysis data as pseudo observations in locations of the reanalysis grid. In addition, the pseudo observations could be chosen for example, to represent only large-scale atmospheric features (e.g., applying a low-pass filter to the data).

[15] We wish to point out that the linear equation (3a) does not imply that the response of the state to the forcing is assumed linear. The linear equation given by (3a) includes the relationship between the variations in the state and forcing by the cross-covariance (equations (A3a) and (A3b)). The cross-covariance in the EnKF formulation captures the nonlinear interactions between the state and forcing because it is derived from the ensemble of fully nonlinear model simulations. This is a well known property of the EnKF estimates [*Evensen*, 2006]. This property implies that the estimates of $x_{\text{no forcing}}$ by the IFPRE method would include the nonlinear interactions, but would correspond to the statistical mean state instead of the state which would result from deterministic nonlinear evolution of the true atmosphere with the forcing being exactly zero.

[16] The relationship between the mean estimate and the model nonlinearity is well known from the statistical estimation theory. The theory shows that the mean estimate does not represent the state which would result from the deterministic nonlinear response of the state to the mean forcing [*Tarantola*, 2005]. The discrepancy between the mean estimate and the deterministic nonlinear response could be large if the nonlinearity is strong [*Vukicevic and Posselt*, 2008]. Consequently, if the model has significant nonlinearities, the estimate by the IFPRE method, equivalently to the estimate by the EnKF method, would not satisfy exactly the deterministic nonlinear relationship between the assumed statistical mean forcing and the state. However the statistical mean estimate would be a desired estimate because of the uncertainties in the observationally based estimate of forcing, the model and the analysis data. Examples of the influence of the nonlinear relationship between the state and forcing on the mean state estimates are shown in the next section. Ideally, the estimates of state with or without the forcing should be presented in terms of multidimensional and multivariate probability density function which would not be Gaussian in the presence of the nonlinear relationship between the forcing and state. Such presentation is unattainable in reality because of the large number of degrees of freedom in the atmospheric system

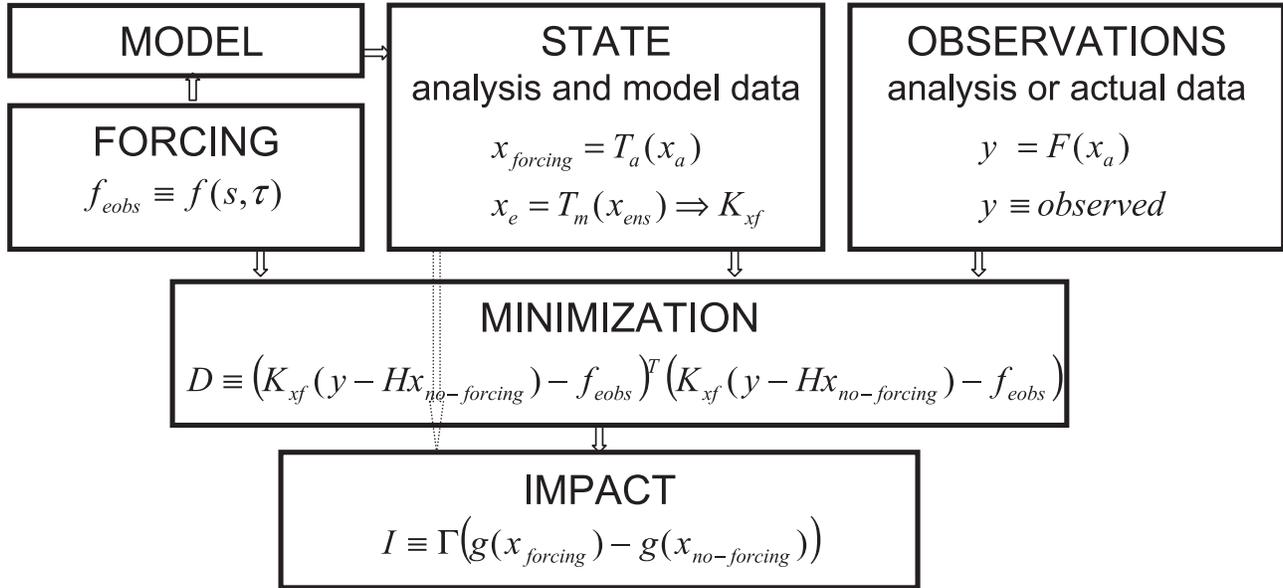


Figure 1. Schematic of the impact of forcing by prior state estimating (IFPRE) method. Arrows indicate data flow. The symbols T_a and T_m in the “STATE” panel denote transformation operators, respectively, from the standard reanalysis and model data into the state grid in the estimation problem. The symbol F in the “OBSERVATIONS” panel denotes transformation from the standard reanalysis data into the observation space. This transformation produces pseudo observations. Other symbols are defined in the text.

and because the exact, possibly non-Gaussian, statistics of the data uncertainties are not known.

[17] To summarize, the new method, referred as IFPRE, for the assessment of the impact of the heterogeneous forcing on atmospheric circulation would include the following procedures, also displayed in Figure 1.

[18] 1. Specification of the state vectors ($x_{no-forcing}$ and $x_{forcing}$), involving selection of the temporal and spatial scales of interest and the state variables in the impact assessment problem and transformation of the model ensemble and reference atmospheric analysis (e.g., the reanalysis) data into the state vectors (panel labeled STATE in Figure 1). For example, if the interest is in the impact of aerosol radiative forcing on the large-scale circulation within a geographical region such as Southeast Asia as in the work of Kim *et al.* [2007] and Ramanathan *et al.* [2005], then the state could be represented by large-scale geopotential height fields from reanalysis and model results within the region. Assuming additionally that the main interest is in monthly mean response to the forcing, then the model ensemble and reanalysis data would be monthly averaged. With such definition of the state, the IFPRE analysis would address the problem of the impact of the aerosol forcing on monthly mean large-scale circulation as represented by the geopotential fields in the Asia region. Equivalently, the state definition could include longer time scales.

[19] 2. Selection of the observations y to use for the estimate of x_{eprior} (panel labeled “OBSERVATIONS” in Figure 1). The role of the observations in estimating the state without the forcing is equivalent to the role the observations have in the standard atmospheric data analysis: to provide the best reference data about the actual atmo-

sphere relative to which the model solution is corrected. In the previously mentioned example, the reference observations data would be from the same region (i.e., Southeast Asia) and should represent the same temporal and spatial scales as in the definition of the state, but the observation vector would include other standard atmospheric variables.

[20] 3. Specification of the observationally based forcing (f_{eobs}) in the form of heterogeneous forcing function that would be used by the model (panel labeled “FORCING” in Figure 1). The need for the observationally based forcing function in the model implies that the actual observationally based data of the forcing may need to be transformed into the model defined representation of the forcing. The form of this transformation would depend on the model and the original observationally based forcing data.

[21] 4. Solving the least square problem (4) for the estimate of the statistical mean atmospheric state unaffected by the forcing, x_{eprior} . This procedure requires use of a numerical minimization algorithm suitable for linear problems, usually readily available in standard numerical libraries.

[22] 5. Evaluation of the forcing impact using equation (1) with the suitable diagnostic function and measure of distance. In the hypothetical example of the regional impact on the large-scale monthly mean geopotential fields, the suitable diagnostic function could be simply $g(x) = x$, because the spatial filtering and time averaging are included in the definition of the state. The distance operator $\Gamma(g(x))$ could include longer time averages (e.g., the season) and, for example, absolute distance or quadratic norm.

[23] In the next section we demonstrate the new method by a tutorial example using the large-scale circulation model of Lorenz [1984, 1990]. This model was also used in Pielke

and Zeng's [1994] theoretical study of long-term variability of the atmosphere under the influence of short-term variable forcing.

3. Demonstration of the IFPRE Method by a Tutorial Example

[24] The large-scale circulation parameterization is described by Lorenz [1984, 1990] and Pielke and Zeng [1994]. We present here only a brief description of the model pertinent to the estimation experiments by the IFPRE method. The model governing equations are

$$\begin{aligned}\frac{dX}{dt} &= -Y^2 - Z^2 - aX + aF \\ \frac{dY}{dt} &= XY - bXZ - Y + G \\ \frac{dZ}{dt} &= bXY + XZ - Z,\end{aligned}\quad (5)$$

where, X represents the strength of large-scale westerly current, and Y and Z represent the superimposed waves with cosine and sine phase, respectively. The dynamical system described by (5) is a dissipative forced system. The forcing is represented by symmetric (G) and asymmetric (F) thermal forcing contributions.

[25] Similar to the Lorenz [1990] and Pielke and Zeng [1994], in this study the values of $b = 4$, $a = 0.25$ and $G = 1$ are used together with the initial conditions $X(0) = 2$, $Y(0) = 1$ and $Z(0) = 0$, and time step of 0.025 units in the fourth-order Runge Kutta solver. In the current numerical experiments, the time invariant asymmetric forcing F is used to exemplify the conditions of the model with and without the forcing. For this purpose the asymmetric thermal forcing is expressed as

$$F = F_0 + f, \quad (6)$$

where, F_0 represents a base thermal forcing and f is used to simulate the equivalent of the heterogeneous forcing. Thus, only f was varied between the model versions with and without forcing in the estimation experiments. Relatively small variations in the forcing of $\Delta f = 2$ units were used to represent the change between forcing and no-forcing conditions because the variation was intended to exemplify an incremental change in the atmospheric heterogeneous forcing and because the model in (5) is highly and nonlinearly sensitive to the change in forcing. These attributes make the current model suitable for testing the influence of nonlinearities on the estimates by the new method. As discussed in section 2, the new method makes use of the statistical mean estimate of the state. This estimate would by definition differ from the deterministic nonlinear response to the assumed change in the forcing. In the case of a perfect model the deterministic nonlinear response is the true response. Consequently, in the tutorial example presented here the discrepancies between the deterministic nonlinear and the mean state estimates are evaluated to demonstrate the skill of the method.

[26] In the design of the experiments it was considered that the degree of nonlinearity depends on the model dynamics and time scale. The time scale for testing the

impact of the forcing is by definition bounded by the predictability time scale which, as is well known, depends on what aspect of the prediction is considered. For example, the predictability time limit for the large-scale monthly mean circulation is longer than the equivalent for the local mesoscale circulation. In the current experiments with the idealized and relatively simple model of the midlatitude atmospheric dynamics the prediction of instantaneous and weekly average states are considered. The aspect of prediction in the estimate of the forcing impact by the IFPRE method corresponds to the procedure of state definition (panel labeled "STATE" in Figure 1).

[27] Several estimation experiments were performed with different values of total asymmetric forcing to test the sensitivity of the method to the temporal variability in the model solution. The temporal variability of the state (X , Y , Z) is strongly driven by the amplitude of the asymmetric thermal forcing and nonlinear interactions between the state components. A minimum value of $F = 3$ was used in the estimation experiments because the model is highly dissipative for a small amount of the forcing; when $F < 3$ the solution asymptotes to a constant value after less than the equivalent of several weeks. The experiments consisted of numerically solving equation (4) for the state without the forcing ($f = 0$) using two different definitions of the state as follows: (1) instantaneous values of (X , Y , Z), 12 h apart within a period of one month (equivalent of 500 model time steps) and (2) time average of (X , Y , Z) over 7 day periods with the lag of 12 h within the same monthly period. For each definition of the state the data needed for the estimate of impact were produced in the following way:

[28] 1. The equivalent of observation data were produced by the model with $f = 2$, over a period of 1 month. The simulated observations were distributed at even time intervals of 12 h, and the error, with a mean of zero and standard deviation of one, was added to the data to simulate the observation errors.

[29] 2. The equivalent of the reanalysis data was produced by an independent EnKF data assimilation algorithm using the observations of only the X component of the model state and the model with $f = 0$. Specifically, the data assimilation algorithm was based on the ensemble square root filter (EnSRF) approach after Whitaker and Hamill [2002]. In the EnSRF assimilation the initial condition was perturbed, instead of the forcing, to produce the ensemble error covariances in the state space. The data assimilation was performed sequentially using the analysis window of 12 h. The simulated observations and states produced by the EnSRF analysis are shown in Figure 2 for the model version with $F = 8$. As expected, the EnSRF analysis does not completely correlate with the observations except for the X component which was assimilated.

[30] 3. The ensemble data for computing the cross-covariance between the state and forcing were produced by 10^3 model simulations using the values of f from the normal distribution with a mean of 2 and standard deviation of 0.5. Because the state in the current experiments is the sequence of either the instantaneous model solution or the running weekly average, the ensemble simulations were performed sequentially using the time window of 7 days starting every 12 h. This procedure produced 57 ensembles with 10^3 members each. The period of 7 days was chosen as the

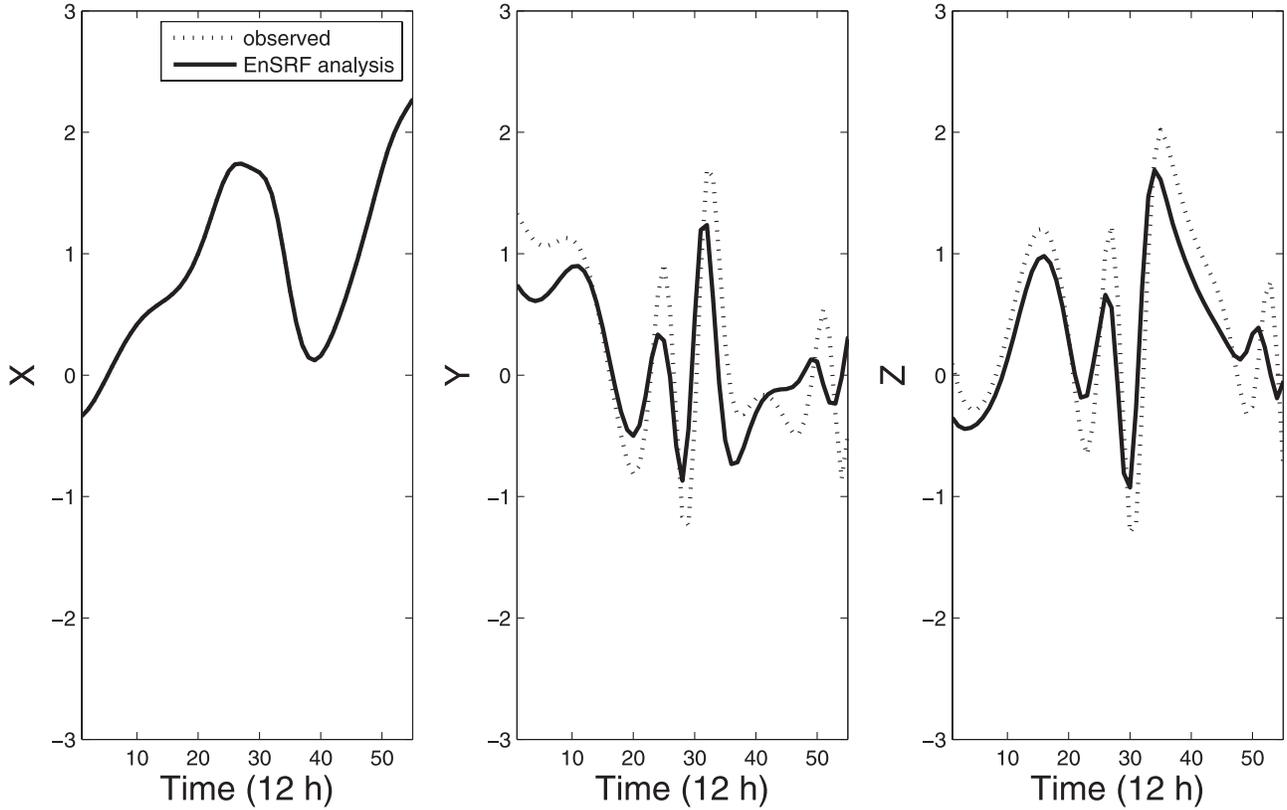


Figure 2. State components (X , Y , and Z) of Lorenz's large circulation model as a function of time with the asymmetric thermal forcing $F = 8$ (time unit on the horizontal axis is 12 h). Solid curves are for the state analysis by the ensemble square root filter (EnSRF) data assimilation, and dotted curves are for the simulated observations.

predictability time limit for the instantaneous state in the current model.

[31] We present the results of estimation experiments for two asymmetric thermal forcing scenarios. The scenarios were produced with the asymmetric thermal forcing values of $F = 5$ and $F = 8$ which were associated, respectively, with the lower- and higher-frequency evolution of the state (X , Y , Z). For both forcing scenarios the estimates of state with $f = 0$ (i.e., the state equivalent to missing the heterogeneous forcing) were computed, implying that the base state F_0 in equation (6) was different between the experiments. In all experiments the theoretical minimum value of $D = 0$ in equation (4) was achieved. The estimates of the impact of forcing were computed using the following simple diagnostic function and measure of distance, respectively: $g(x) = x$ and $\Gamma(g) = g_{\text{no forcing}} - g_{\text{forcing}}$. The ideal or *true* values of x_{forcing} were obtained from the simulated observations without the observation errors. Equivalently, the model simulation with $f = 0$ at 12-h intervals was used as the true state without the forcing. The estimates of the forcing impact were computed using the following expressions:

$$I_{\text{true}} = g(x_{\text{forcing}}^t) - g(x_{\text{no forcing}}^t), \quad (7a)$$

$$I_1 = g(x_{\text{forcing}}^t) - g(x_{\text{no forcing}}^e), \quad (7b)$$

$$I_2 = g(x_{\text{forcing}}^a) - g(x_{\text{no forcing}}^e), \quad (7c)$$

where, x_{forcing}^t , x_{forcing}^a , $x_{\text{no forcing}}^t$, and $x_{\text{no forcing}}^e$ are, respectively, the true and analyzed states with the forcing and the true and estimated states without the forcing. The analyzed state is the result of the data assimilation by the EnSRF algorithm.

[32] The estimates of the forcing impact by expressions (7b)–(7c) for the instantaneous states are compared to the true impact (I_{true}) in Figures 3 and 4. The results show that the estimates using the true reference (the expression 7b and dotted curves in the Figures 3 and 4) are in close agreement with the true deterministic solution except during periods of rapid phase change in the state. During these periods the amplitude of impact is over or under estimated with an average error of about 20%. Because the temporal variability is smaller in the experiment with $F = 5$ (Figure 3) than with $F = 8$ (Figure 4) the results are overall slightly better with the smaller forcing. The comparison between the estimates of impact relative to the true and the analyzed state (the dotted and dashed curves in Figures 3 and 4) shows that the estimates tend to be closer to the deterministic values when the true state is used for the reference. This result is expected because of the errors in the EnKF analysis relative to the true state. In reality, the true state of

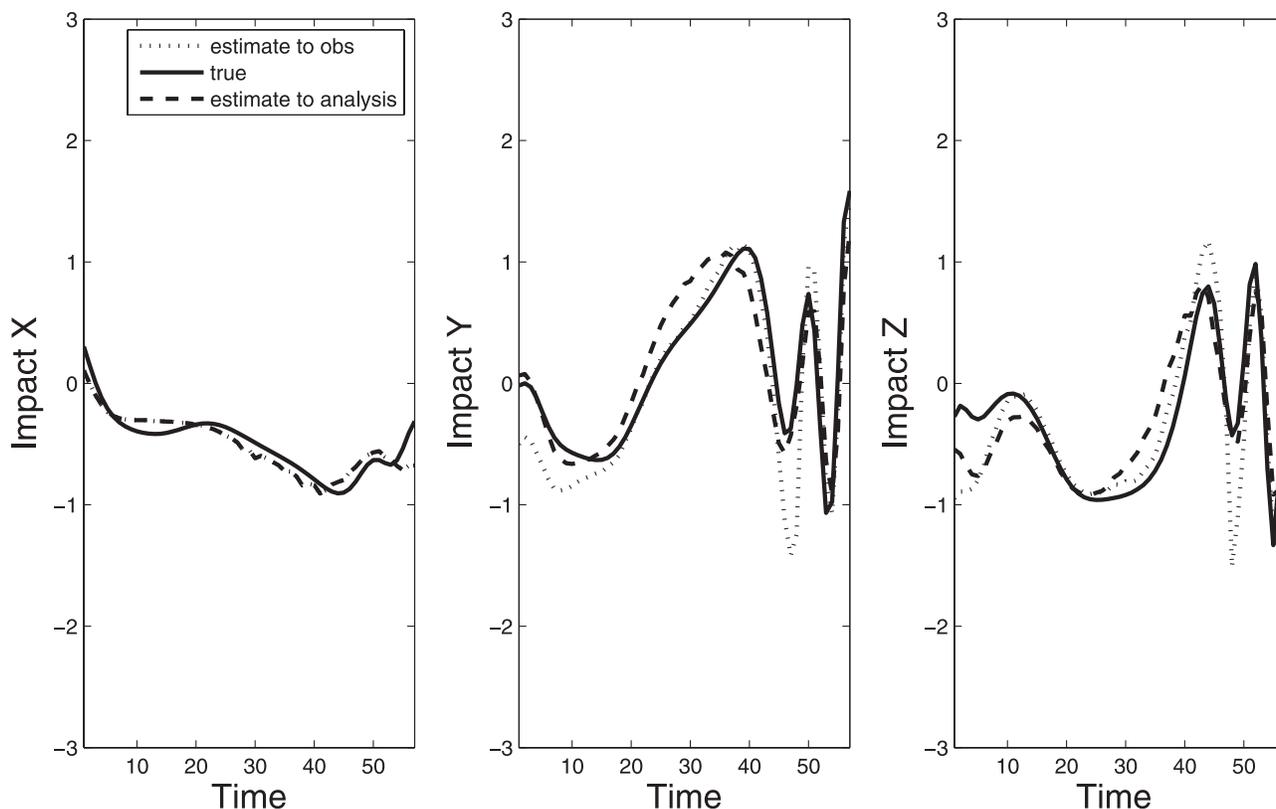


Figure 3. The impact of change in the asymmetric thermal forcing on the instantaneous state at different times in the experiment with $F = 5$. The true deterministic impact is displayed by the solid curves. The estimates of the impact by the IFPRE method are displayed by the dashed curves for the reference data by the EnSRF analysis and by the dotted curves for the true reference data. The panels from left to right are for the state components X , Y , and Z , respectively. The time unit is 12 h on the horizontal axis.

the actual atmosphere is not known, implying that the uncertainties in the atmospheric analysis data would influence the estimate of impact by the new method. We wish to point out that the uncertainties in the state estimates by the standard data assimilation are by definition smaller than the uncertainties in the model simulations. This property implies that the analysis of the forcing impact by the new method would be less erroneous than the standard model-based sensitivity analysis. The impact of the model errors is not explicitly evaluated in the current examples because of the use of the idealized model.

[33] The estimates of the forcing impact for the 7-day averaged states are compared to the equivalent true deterministic impact in Figures 5 and 6. The results are similar to the experiments with the instantaneous state except that the smaller temporal variability in the state produces slightly better agreement between the true and estimated impact. Additional experiments were conducted to test the sensitivity of the estimates to specification of the observation vector in equation (4). As pointed out in section 2, the observation vector may include the pseudo observations from the reanalysis data instead of the actual observations. This approach was tested in the current experiments by replacing the simulated observations with the equivalent from the results of the EnSRF data assimilation. The relatively small change in observation values produced negligible differ-

ences in the estimates by the IFPRE method (not shown). These results, although limited to a relatively simple model and idealized conditions, suggest that the new method could be applied with the pseudo observations.

[34] The current tutorial examples demonstrate that the IFPRE method produces the skilled estimates of the forcing impact within limits of the estimation approach which is based on the statistical mean estimates and the Gaussian error statistics. The mean estimates would differ from the true deterministic impact in the presence of nonlinear relationship between the forcing and atmospheric state. Given that the true state cannot be known because the models and observationally based data of the atmosphere are inherently erroneous and that the full non-Gaussian probability density cannot be evaluated for the complex system with large number of degrees of freedom, the analysis of the forcing impact by the mean estimates is the optimal approach.

4. Summary and Conclusions

[35] In this study we present a new method for estimating the impact of heterogeneous forcing on atmospheric circulations designated IFPRE. This new method is similar to the commonly used model-based sensitivity studies in that the impact of forcing is diagnosed by a suitable measure of

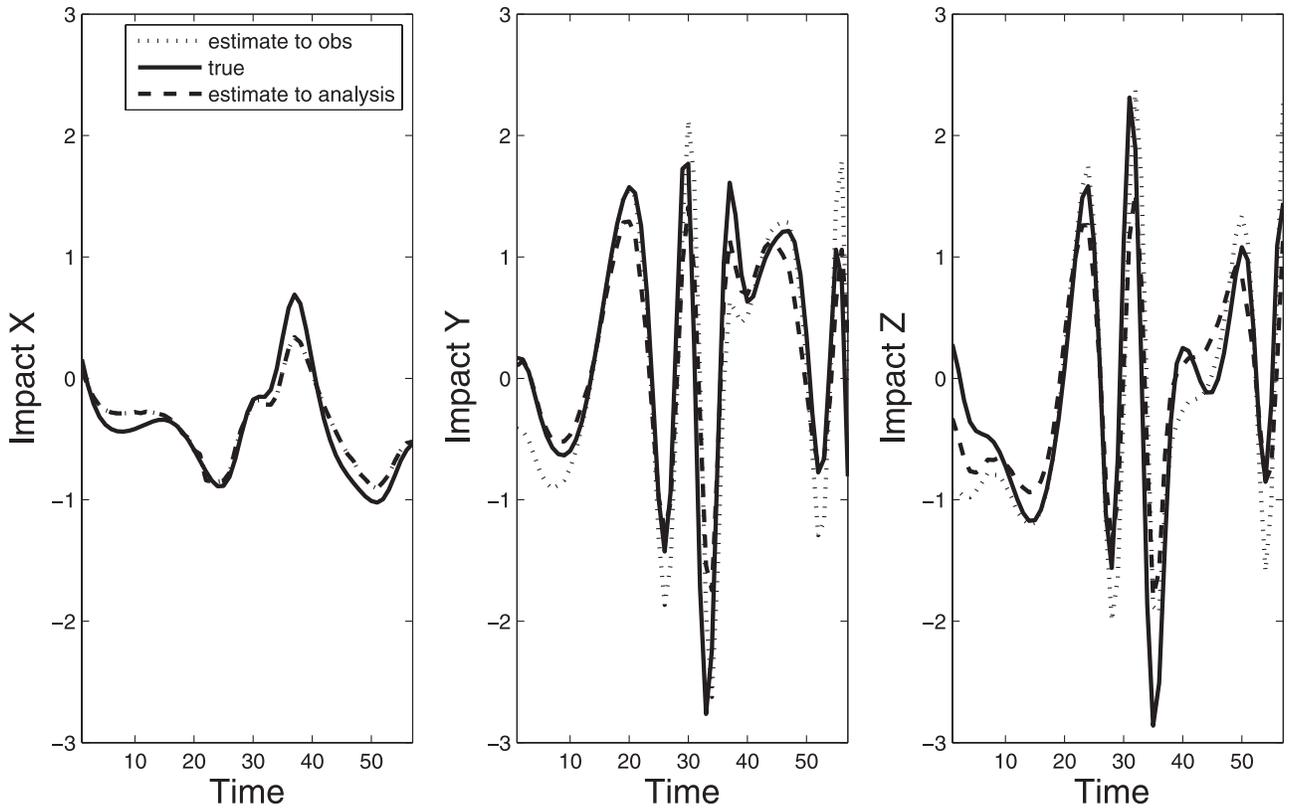


Figure 4. Same as Figure 3, but for the experiments with $F = 8$.

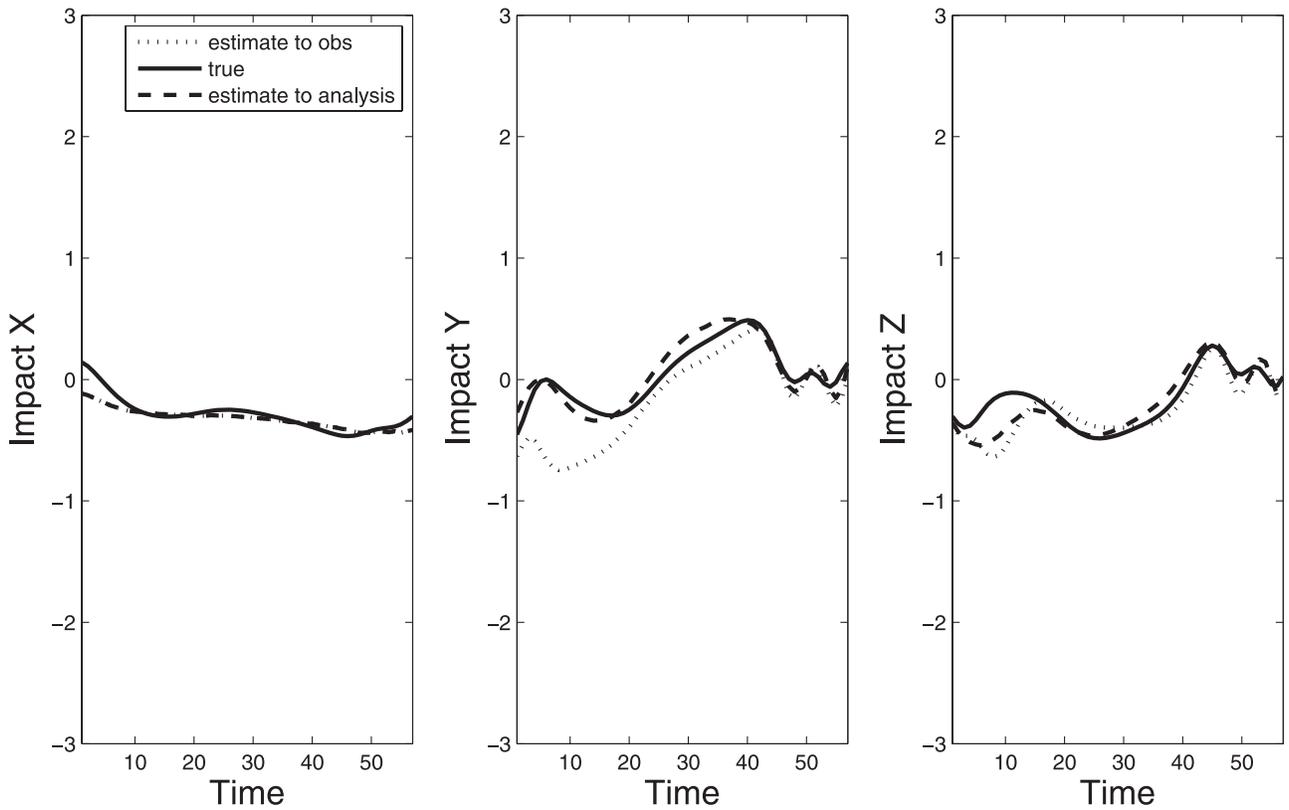


Figure 5. Same as Figure 3, but for the 7-day averaged state.

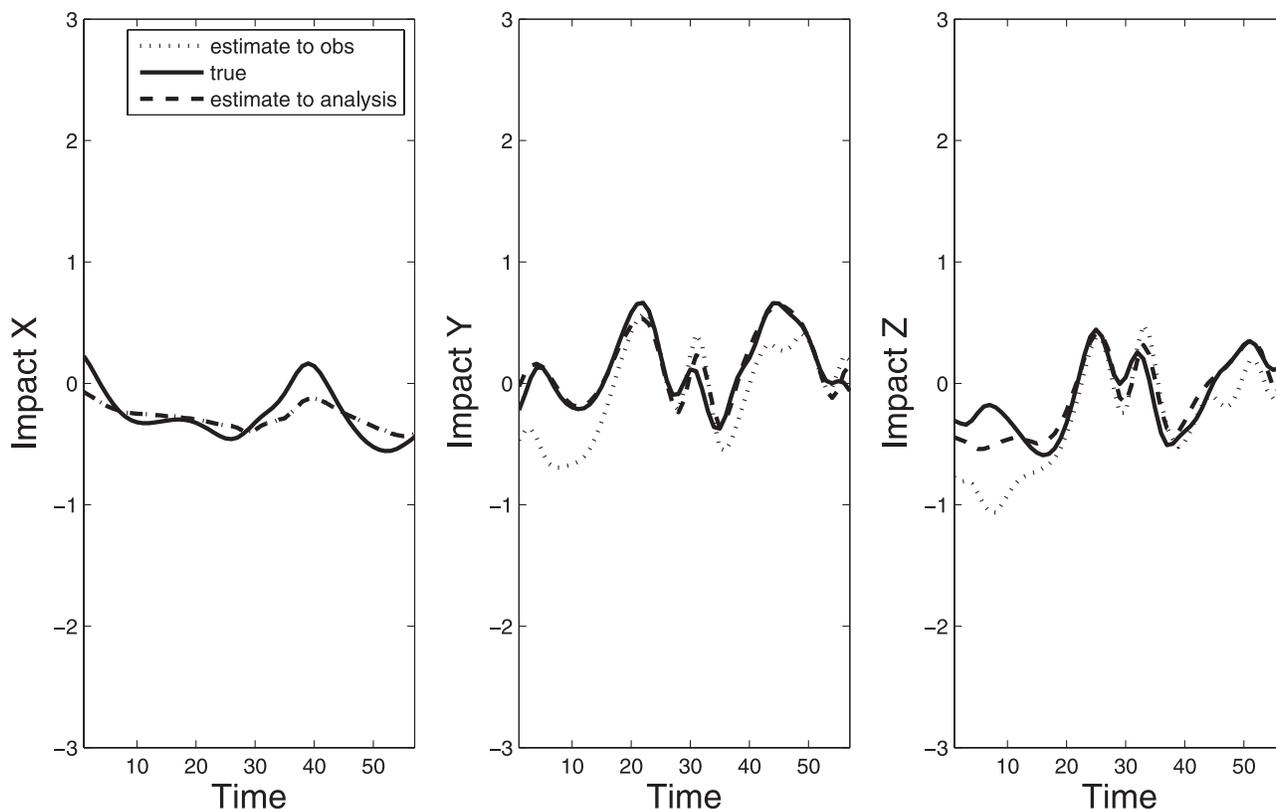


Figure 6. Same as Figure 4, but for the 7-day averaged state.

difference between the atmospheric states with and without forcing, but differs from it in the way these atmospheric states are evaluated. In the IFPRE method the atmospheric states with and without the forcing are evaluated using the data assimilation approach instead of using only model simulations. The main feature of the IFPRE method is that the estimates of the atmosphere with and without the influence of forcing are observationally based. The new method includes the following procedures:

[36] 1. The estimate of atmospheric state with forcing is obtained from standard atmospheric reanalysis data.

[37] 2. The estimate of atmospheric state without forcing is obtained by the numerical solution of the least square derived from the ensemble Kalman filter data assimilation approach (section 2).

[38] 3. The impact of forcing is evaluated using the relationship

$$I = \Gamma(g(x_{\text{forcing}}) - g(x_{\text{no forcing}}))$$

where, as before, Γ and $g(x)$ are, respectively, a suitable measure of distance and diagnostic of the atmospheric states with forcing (x_{forcing}) and without ($x_{\text{no forcing}}$).

[39] In this study the IFPRE method is demonstrated in a tutorial manner using the large-scale circulation model of Lorenz [1984, 1990]. In the numerical experiments with this model, synthetic observation data were obtained by the model simulations using standard values of the parameters as in the Lorenz [1990] study. The equivalent of atmospheric reanalysis data were produced using the EnSRF algorithm

with the simulated observations of only one of three components of the model state. Variations in the parameterized asymmetric thermal forcing in the model were used to simulate the heterogeneous forcing mechanism. The results with two different dynamical regimes in the state evolution and different specification of the observation data demonstrate that the IFPRE method is capable of producing skilled estimates of the mean forcing impact within the predictable time limits.

[40] The key feature of the method is objective estimation of the atmospheric state that is unaffected by the forcing. The main practical issues in this estimation problem are related to (1) ensemble simulations with a GCM, which would include the observationally based estimate of the temporal and spatial distribution of the forcing, and (2) the solving of a potentially large minimization problem.

[41] Regarding the application of the new method with real world data and models, and assuming that the observationally based estimates of the heterogeneous forcing are available (e.g., as summarized by Kabat *et al.* [2004] and NRC [2005]), the ensemble simulations would require the capability to include forcing in the GCM with the same spatial and temporal distribution as in the observation estimates. This could be achieved by replacing the model parameterization with a look-up table of the forcing data [Pielke *et al.*, 2007]. The look-up table would also be used to generate ensemble perturbations in the forcing. The GCM ensemble simulations using the observationally based estimates of the forcing distribution would produce valuable data not only for estimation by the new method but also for

more traditional studies, such as validation of parameterizations in the model that are not based on just observations, the sensitivity of the impact assessment to the model uncertainties, and the testing of probabilistic measures of the model projections using the ensemble data.

[42] The minimization problem given by equation (4) in section 2 could be too large to solve if the atmospheric state included all of the standard variables and full grid resolution of the reanalysis or the GCM, and if the observation vector would include the global standard observing network data. The primary computational challenge is in evaluating the Kalman gain matrix which involves inversion of the matrix A that is of the size of square of the number of observations (expression (A3b)). The problem would become computationally tractable if the effective observation spaces would be reduced.

[43] An effective and physically justifiable reduction of the space could be achieved by considering that the primary interest regarding the impact of heterogeneous forcing is to evaluate changes in the synoptic to large-scale circulation patterns; globally or regionally. The focus on the larger scales or regional impact suggests that the estimation problem which includes the solving of minimization problem together with evaluation of the Kalman gain matrix should be defined from the start in terms of the appropriate scales and domain. In this way the dimension of the state and observation spaces would be significantly reduced. The scale definition of the observation and state spaces could be obtained either by spatial averaging or other standard filtering approach on the data used in the problem (i.e., the model ensemble, reanalysis and observation data). In addition, the pseudo observations from the reanalysis data could be used instead of the actual observations for more effective scale filtering in the observation space. The sensitivity of the estimates to the source of observation data should be tested in the applications.

[44] In summary, we wish to emphasize that a strong potential for more accurate and more robust estimates of the impact of forcing by the new method which would result from the utilization of the observationally based data of the atmosphere and forcing and from using the statistical mean estimates should motivate resolving of the computational challenges. An additional appealing property of the method is that reducing of the errors in the observationally based data would improve the estimates of the impact of forcing automatically unlike in the model sensitivity studies which are largely independent of advances in the observing and data assimilation. The computational challenges of the IFPRE method with the real world data and models are in the domain of difficulties which could be reduced by technical means and clever design.

Appendix A: Ensemble Kalman Filter Data Assimilation

[45] The EnKF class of data assimilation techniques originate from standard statistical estimation theory and uses probabilistic representation of information in the data about a simultaneously modeled and observed dynamical system [Tarantola, 2005; Jazwinski, 1970]. In the case of atmospheric data assimilation the data consists of the atmospheric observations and forecast model results. The

result of assimilation is gridded atmospheric analysis. By the statistical estimation theory the data assimilation problem is written

$$p(x/y) \approx \exp(-J(x)) \quad (\text{A1a})$$

$$J(x) = \frac{1}{2} \left[(h(x_{\text{prior}}) - y)^T R^{-1} (h(x_{\text{prior}}) - y) + (x - x_{\text{prior}})^T P^{-1} (x - x_{\text{prior}}) \right], \quad (\text{A1b})$$

where $p(x/y)$ is the posterior probability density function (pdf) of the atmospheric state (x) on an analysis grid, given the observations (y). J is the so-called cost function which includes two quadratic error measures, defined as follows: (1) weighted difference between the model solution (x_{prior}) mapped to observation space by $h(x_{\text{prior}})$ and the observations (y), where the weights are provided by inverse of observation error covariance R^{-1} , first term in the right-hand side of equation (A1b), and (2) weighted difference between the analysis and the model solution, with the weights given by inverse of prior error covariance P^{-1} (second term in right-hand side of equation (A1b)). The expressions (A1a) and (A1b) are derived using the assumption that the prior (i.e., the model solution) and observation data are stochastic quantities with normally distributed errors with zero mean and covariance P^{-1} and R^{-1} , respectively.

[46] Using the EnKF technique, the data assimilation problem (A1) is solved by numerically estimating mean and covariance of $p(x/y)$ (i.e., the pdf of gridded analysis). The data assimilation solution is written

$$x_e = x_{\text{prior}} + K(y - Hx_{\text{prior}}), \quad (\text{A2a})$$

$$P_e = (H^T R^{-1} H + P_{\text{ens}}^{-1})^{-1}, \quad (\text{A2b})$$

$$K = P_{\text{ens}} H^T (H P_{\text{ens}} H^T + R)^{-1}, \quad (\text{A2c})$$

where, x_e and P_e are estimates of posterior mean and covariance, respectively. The posterior mean is the optimal analysis estimate, while the covariance represents the associated uncertainty. The matrix P_{ens} is an ensemble estimate of the prior error covariance, where the ensemble consists of a sample of different model solutions. H is the linearized version of the mapping function $h(x)$ which is used in expression (A1b). The matrix K is typically referred to as Kalman gain matrix, after *Kalman* [1960], who first introduced the Kalman Filter estimation approach. The EnKF techniques differ from each other primarily in ways the matrices P_{ens} and K are evaluated from the ensemble of model solutions and by how the ensemble data are generated [Evensen, 2006]. In practice the EnKF solution (A2a)–(A2c) is used sequentially over time at discrete time instances and is referred to as the analysis time. At every analysis time, a new set of observations is used which corresponds to a time window around that time. The skill of the analysis by the EnKF techniques have been demon-

strated in many applications including atmospheric, oceanic, land surface, soil, and hydrologic data analysis [Miyoshi and Yamane, 2007; Zhang and Snyder, 2007; Evensen, 2006; Leeuwenburgh, 2007; Montaldo et al., 2007].

[47] The EnKF technique has also been applied in climate studies to estimate optimal values of parameters in climate models and to evaluate climate variability by observationally constrained model simulations [Annan et al., 2005; Aksoy et al., 2006; Hargreaves and Annan, 2006]. The parameter estimation was employed to objectively improve the model skill in critical aspects of the modeling such as parameterization of radiation, clouds, and precipitation [Annan et al., 2005]. The parameter estimation and the associated evaluation of uncertainty in modeling and prediction by the ensemble data assimilation has been recognized as a promising new approach in climate research [Randall et al., 2007].

[48] The data assimilation solution when both the state and model parameters are being estimated is similar to the solution in (A2) but the modeled and analyzed state quantities x and x_e , respectively, are augmented to include the parameters. The mean solution is then written

$$\begin{pmatrix} x_e \\ \alpha_e \end{pmatrix} = \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix} + \begin{bmatrix} K_x \\ K_{x\alpha} \end{bmatrix} \left(y - [H \ 0] \cdot \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix} \right), \quad (\text{A3a})$$

$$\begin{bmatrix} K_x \\ K_{x\alpha} \end{bmatrix} = \begin{bmatrix} P_{\text{ens}} & P_{x\alpha} \\ P_{x\alpha} & P_{\alpha} \end{bmatrix} [H^T 0] A^{-1}, \quad (\text{A3b})$$

$$A \equiv [H \ 0] \begin{bmatrix} P_{\text{ens}} & P_{x\alpha} \\ P_{x\alpha} & P_{\alpha} \end{bmatrix} [H^T 0] + R,$$

where $P_{x\alpha}$ is the cross-covariance between errors of the modeled state x_{prior} and parameters α_{prior} and P_{α} is the autocovariance of the prior parameter errors. All other quantities are defined the same as in (A2). Similar to P_{ens} , $P_{x\alpha}$, and P_{α} are the prior error covariances which are estimated from the model ensemble. The expression

$$[H \ 0] \cdot \begin{pmatrix} x_{\text{prior}} \\ \alpha_{\text{prior}} \end{pmatrix}$$

indicates that only the modeled state is mapped into the observation space as in the solution (A2a).

[49] **Acknowledgments.** During the course of this study, T.V. was supported by NSF grant ATM-0514399. R.A.P. received support through the University of Colorado at Boulder (CIRES/ATOC). A.B.-P. was supported by NSF grants DEB 0217631 and DEB 0080412, NASA grant NNX06AG74G_S02, and NOAA grant NA17RJ1229.

References

Aksoy, A., A. Zhang, and J. W. Nielsen-Gammon (2006), Ensemble-based simultaneous state and parameter estimation with MM5, *Geophys. Res. Lett.*, *33*, L12801, doi:10.1029/2006GL026186.

Annan, J. D., D. J. Lunt, J. C. Hareaves, and P. J. Valdes (2005), Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, *Nonlinear Process. Geophys.*, *12*, 363–371.

Cai, M., and E. Kalnay (2005), Can reanalysis have anthropogenic climate trends without model forcing?, *J. Clim.*, *18*, 1844–1849, doi:10.1175/JCLI3347.1.

Climate Change Science Program (2006), *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*, edited by T. R. Karl, S. J. Hassol, C. D. Miller, and W. L. Murray, Clim. Change Sci. Prog. and Subcomm. on Global Change Res., Washington, D. C.

Evensen, G. (2006), *Data Assimilation: The Ensemble Kalman Filter*, 280 pp., Springer, New York.

Hargreaves, J. C., and J. D. Annan (2006), Using ensemble prediction methods to examine regional climate variation under global warming scenarios, *Ocean Modell.*, *11*, 174–192, doi:10.1016/j.ocemod.2004.12.004.

Intergovernmental Panel on Climate Change (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., Cambridge University Press, New York.

Jazwinski, H. (1970), *Stochastic Processes and Filtering Theory*, *Math. Sci. Eng.*, vol. 64, 376 pp., Elsevier, New York.

Kabat, P., M. Claussen, P. A. Dirmeyer, J. H. C. Gash, L. Bravo de Guenni, M. Meybeck, R. A. Pielke Sr., C. J. Vorosmarty, R. W. A. Hutjes, and S. Lutkemeier (Eds.) (2004), *Vegetation, Water, Humans and the Climate: A New Perspective on an Interactive System*, 566 pp., Springer, Berlin.

Kalman, R. E. (1960), A new approach to linear filtering and prediction theory, *J. Basic Eng., Ser. D*, *82*, 35–45.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, 341 pp., Cambridge University Press, New York.

Kalnay, E., M. Cai, H. Li, and C. J. Tobin (2006), Estimation of the impact of land-use changes and urbanization on climate trends east of the Rockies, *J. Geophys. Res.*, *111*, D06106, doi:10.1029/2005JD006555.

Kim, M.-K., W. K. M. Lau, K.-M. Kim, and W.-S. Lee (2007), A GCM study of effects of radiative forcing of sulfate aerosol on large scale circulation and rainfall in East Asia during boreal spring, *Geophys. Res. Lett.*, *34*, L24701, doi:10.1029/2007GL031683.

Leeuwenburgh, O. (2007), Validation of an EnKF system for OGCM initialization assimilating temperature, salinity, and surface height measurements, *Mon. Weather Rev.*, *135*, 125–139, doi:10.1175/MWR3272.1.

Lewis, J., S. Lakshminarayanan, and S. Dhall (2006), *Dynamic Data Assimilation: A Least Squares Problem*, 680 pp., Cambridge University Press, New York.

Lorenz, E. N. (1984), Irregularity: A fundamental property of the atmosphere, *Tellus, Ser. A*, *36*, 98–110.

Lorenz, E. N. (1990), Can chaos and intransitivity lead to interannual variability?, *Tellus, Ser. A*, *42*, 378–389.

Matsui, T., and R. A. Pielke Sr. (2006), Measurement-based estimation of the spatial gradient of aerosol radiative forcing, *Geophys. Res. Lett.*, *33*, L11813, doi:10.1029/2006GL025974.

Menon, S., J. E. Hansen, L. Nazarenko, and Y. Luo (2002), Climate effects of black carbon aerosols in China and India, *Science*, *297*, 2250–2253, doi:10.1126/science.1075159.

Miyoshi, T., and S. Yamane (2007), Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution, *Mon. Weather Rev.*, *135*, 3841–3861, doi:10.1175/2007MWR1873.1.

Montaldo, N., J. D. Albertson, and M. Mancini (2007), Dynamic calibration with an ensemble Kalman filter based data assimilation approach for root-zone moisture predictions, *J. Hydrometeorol.*, *8*, 910–921, doi:10.1175/JHM582.1.

National Research Council (2005), *Radiative Forcing of Climate Change: Expanding the Concept and Addressing Uncertainties*, 208 pp., National Acad. Press, Washington, D. C.

Pielke, R. A., and X. Zeng (1994), Long-term variability of climate, *J. Atmos. Sci.*, *51*, 155–159, doi:10.1175/1520-0469(1994)051<0155:LTVO>2.0.CO;2.

Pielke, R. A., Sr., et al. (2007), Satellite-based model parameterization of diabatic heating, *Eos Trans. AGU*, *88*(8), 96–97.

Ramanathan, V., C. Chung, D. Kim, T. Bettge, L. Buja, J. T. Kiehl, W. M. Washington, Q. Fu, D. R. Sikka, and M. Wild (2005), Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle, *Proc. Natl. Acad. Sci. U. S. A.*, *102*, 5326–5333, doi:10.1073/pnas.0500656102.

Randall, D. A., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the 4th Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 591–662, Cambridge University Press, Cambridge, New York.

Rial, J., et al. (2004), Nonlinearities, feedbacks and critical thresholds within the Earth's climate system, *Clim. Change*, *65*, 11–38, doi:10.1023/B:CLIM.0000037493.89489.3f.

Strack, J. E., R. A. Pielke Sr., L. T. Steyaert, and R. G. Knox (2008), Sensitivity of June near-surface temperatures and precipitation in the

- eastern United States to historical land cover changes since European settlement, *Water Resour. Res.*, 44, W11401, doi:10.1029/2007WR006546.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, 342 pp., SIAM, Philadelphia, Pa.
- Vukicevic, T., and D. Posselt (2008), Analysis of the impact of model nonlinearities in inverse problem solving, *J. Atmos. Sci.*, 65, 2803–2823.
- Whitaker, J. S., and T. Hamill (2002), Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, 130, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- Zhang, F., and C. Snyder (2007), Ensemble-based data assimilation, *Bull. Am. Meteorol. Soc.*, 88, 565–568, doi:10.1175/BAMS-88-4-565.
-
- A. Beltrán-Przekurat, R. A. Pielke Sr., and T. Vukicevic, Department of Atmospheric and Oceanic Sciences, UCB 311, University of Colorado, Boulder, CO 80309-311, USA. (tomislava.vukicevic@colorado.edu)