DATA ASSIMILATION AND NUMERICAL FORECASTING WITH IMPERFECT MODELS: THE MAPPING PARADIGM

Zoltan Toth and Malaquias Peña[1]

National Centers for Environmental Prediction

[1]SAIC at NCEP

Contribution to a special issue on Data Assimilation in Physica D

26 September, 2005

Revised 15 June 2006

Corresponding author's address: Zoltan Toth, NCEP, Environmental Modeling Center, 5200 Auth Rd., Room 207, Camp Springs, MD 20746

e-mail: Zoltan.Toth@noaa.gov

ABSTRACT

Errors in numerical forecasts arise due to errors in the initial conditions and the discrepancies between the model and nature (and may amplify due to chaos) In a quest to reduce forecast errors, initial conditions for forecast integrations are traditionally chosen to be as close to nature as possible. When such an initial condition (analysis) is used to initialize an imperfect model that is systematically different from nature, the model will drift from a state on or near the attractor of nature to a state near the model's attractor. Such a drift will induce forecast errors.

To reduce drift-induced errors, a mapping paradigm is proposed where a link (i.e., mapping vector) is established between states of nature and corresponding states on (or near) the model attractor. Observations from near the attractor of nature are moved with the mapping vector to the vicinity of the model attractor. Data assimilation is performed with the mapped observations and the mapped initial conditions are then used to initialize model forecasts to be used in the next assimilation cycle. For practical applications, the mapped initial conditions as well as the forecasts are "remapped" back to be close to nature using the mapping vector with an opposite sign.

The mapping paradigm is demonstrated in a setting where a simple Lorenz model is used to generate "nature" and a modified version is used as an imperfect model. The mapping vector is first estimated as the difference between the climate mean of nature and the model. Model-related errors in the Lorenz system with the mapping algorithm are reduced by 67%, leading to improvements in the quality of both the numerical forecasts made with the imperfect model and the analyses produced with the forecasts. Considering that the mapping vector may be a function of phase space location or no long-term climatology for nature or the model may be available, an adaptive approach that can be used with a relatively small amount of data was also introduced and successfully tested.

Keywords: Model drift; Asymptotic model error; Initializing imperfect models; Numerical forecasting

Motto: Reflection is a process where a step back takes you closer to reality

1. Introduction

The quality of numerical forecasts generated by a given system is a function of the quality of the model and the initial condition. According to traditional thinking, forecast errors can be reduced either by increasing the realism of numerical models (i.e., how closely the model resembles nature) or by bringing the initial conditions closer to nature. In chaotic systems like the atmosphere and many other natural systems, the quality of the initial conditions is particularly important since initial errors contribute to a loss of predictability, and the actual lead time at which this loss occurs is determined by the size and structure of initial errors (Lorenz 1963, 1982).

The conventional paradigm for numerical forecasting is based on the above conventional wisdom and consists of the following steps:

1) Estimate the state of nature as truthfully as possible (analysis);
2) Run numerical model forecasts from the analysis field;
3) Statistically assess the systematic error in the numerical forecasts;
4) Remove the estimated systematic error from the forecasts.

Modern data assimilation (see, e.g., Kalnay 2002) combines observational data with either raw or bias-corrected short-range forecasts using analysis-forecast cycles of steps 1-2 or 1-4 above, respectively.

It is argued in this paper that the use of the conventional forecast paradigm, to be called the “fidelity paradigm” for its use of initial fields as truthful to nature as possible, is justified only in the special case when perfect models are used for prediction. Such situations arise only in controlled scientific experiments. In the general case when predictions are made with imperfect models (and all practical applications fall into this category) the generally accepted fidelity paradigm may not be valid.

As a simplified example, consider a numerical model as “nature” and a modified version of it as a tool used for predicting nature (i.e., the imperfect “numerical model” of “nature”.) The attractors of these two systems are generally different. When an imperfect numerical model is started from an exact state of nature (i.e., no initial error in the traditional sense, see Fig. 1) the forecast will diverge from nature. This divergence is a transient behavior and is exclusively due to model imperfection (i.e., the difference between nature and its model) and is often referred to as model drift (see, e.g., Toth and Vannitsem 2002).

The working hypothesis of this study is that the initial drift is associated with the choice of an initial state that is on or near the attractor of nature. Such states are obviously not on the attractor of an imperfect model and will necessarily lead to a transient drift. It is further assumed that drift, if present in a numerical forecast, will induce additional forecast errors that would not emerge had an appropriate initial state on or near the model attractor been chosen.

The main goal of this paper is to explore whether drift-induced forecast errors can be reduced by “mapping” an initial state from nature close to the attractor of a different though similar system (i.e., a numerical model of nature). Mapping is defined here as an operation that links each natural state with a state near the model attractor, when used as an initial condition leads to a model forecast that best reproduces the time evolution of nature. Once a forecast is made from a mapped initial condition, it can be remapped back from being near the model attractor to being near that of nature by using the mapping vector with an opposite sign (Fig. 2). The expectation is that if such a mapping can be established and used for initializing numerical forecasts, errors induced by model drift can potentially be reduced, leading to better forecast performance.

The application of the mapping paradigm for data assimilation and numerical forecasting starts with an estimation of the mapping vector connecting the attractors of nature and the model. This is followed by the steps below:

a) Map the observations from the vicinity of nature to that of the model attractor;
b) Assimilate the mapped observations;
c) Run the model from the mapped initial condition;
d) Remap the analysis and forecast back to the phase space of nature (Fig. 3).

The “fidelity” paradigm assumes that nature and the forecast model are the same and, therefore, starts the model with a state as close to the state of nature as possible. This assumption is true only in case of a perfect model (that exists only in simulated numerical experiments.) By contrast, the “mapping” paradigm recognizes that numerical models of any system in nature are only imperfect images of reality and searches for a mapping that connects the states observed with the corresponding states in a model. The mapping is then used to find the model state that best represents a state of nature on or near the model attractor for model initialization.

Tacitly assuming that numerical models are perfect, the traditional paradigm strives at finding initial conditions that match nature as truthfully as possible. In contrast, the mapping paradigm explicitly recognizes the systematic difference between corresponding states of nature and a numerical model of it (i.e., the mapping vector) and considers traditionally derive initial states as, in excess of random type of initial errors, also burdened with an additional error that at initial time equals to the mapping vector. This initial error (i.e., the mapping vector) is the asymptotic (i.e., lead time independent) systematic difference between nature and its numerical model.

Data assimilation cycles with the mapping paradigm consist of steps a-c above. Step d is used outside the cycles of data assimilation, and only when the results are to be used or evaluated in practical applications. And since modern data assimilation (DA) methods rely on short-range forecasts to propagate observational information in a dynamically consistent way in space and time, if mapping can reduce forecast errors, possibly it can also improve the fidelity of the analysis to nature.

After a brief historical overview (Section 2), the numerical model and data assimilation tools used in this study will be presented in Section 3. The mapping paradigm will be demonstrated with a simple Lorenz model, using two different methods for the estimation of the mapping vector in Sections 4 and 5, respectively. Preliminary conclusions and a discussion are offered in Sections 6 and 7, respectively.

2. Historical overview

Accurately reproducing nature in numerical analyses has been a top priority for data assimilation research for decades. Considerable efforts have been made to bring the analysis fields as close to nature as possible. This included not only reducing random errors in analysis fields that are due to observational or other noise, but also efforts to eliminate systematic errors arising due to the use of imperfect models for generating short-range forecasts. Traditionally, model-related systematic errors have been dealt with one of the following two approaches. First, systematic errors in short-range forecasts used as background fields can be estimated, and then removed (see, e.g., Dee and Da Silva 1998, Dee and Todling 2000 and references therein). Alternatively, a special “model error“ covariance matrix can be introduced in data assimilation algorithms. This allows the analysis to deviate from the background field and resemble more the observational data in directions assumed to be associated with model-related errors (see, e.g. Zupanski and Zupanski 2006). In either case the analysis is systematically moved from states that the model prefers toward nature.

While the reduction of random errors in initial fields is certainly necessary in reducing the growth of forecast errors in chaotic systems, fighting the emergence of systematic differences between nature and the assimilating model (and thus the analysis fields) as usually done in accord with the fidelity paradigm can, according to the “mapping” paradigm, be counterproductive. While the mapping paradigm, to our knowledge, has not yet been presented in the literature in a comprehensive way,a few attempts to deviate from the prevailing “fidelity” paradigm must be noted.

First we refer to research related to what is called representativeness errors. Numerical models, even if otherwise very similar to nature, cannot represent features whose spatial and/or temporal scales are finer than the resolution of the model. Atmospheric observations from a convective system with a scale of 1 km, for example, may be very misleading with respect to the state of the flow on the larger meso- (10 km) or synoptic (100 km) scales that a model can represent.

It has been recognized that small-scale features that cannot be resolved by the assimilating model act as noise and will degrade the quality of DA and forecast products. Typically, DA schemes partially address this problem by increasing the value of observational errors in the algorithms so they statistically also account for errors associated with model representativeness (e.g., Lorenc 1986). The desire to exclude features from the analysis that are on scales that the assimilating model cannot resolve deviates from the fidelity paradigm, and is in line with the principles of the mapping paradigm.

In a stimulating study, Schneider et al., (1999) explored components of the mapping paradigm in their experiments with a coupled ocean-atmosphere model. Ocean initial conditions were first derived using the fidelity paradigm. These ocean initial conditions were then mapped onto the phase space of an ocean model coupled with an atmospheric model. This was done by taking the anomalies of the ocean analyses from their long-term climatology and putting these anomalies on the long-term climatology of the ocean component of the coupled ocean-atmosphere model used. In their study, Schneider et al. (1999) reported some modest success with their approach. This is despite some technical issues that could have negatively affected their study, e.g., mapping the conventionally made analysis fields instead of mapping and then analyzing the observations; use of relatively small samples to estimate the climate mean of the analyses (used as a proxy for nature) and the model; lack of mapping for the initialization of the atmospheric component of the coupled ocean-atmosphere model; lack of re-mapping the forecasts into the space of nature (or the traditional analyses); and the use of a relatively short evaluation period.

Another study of interest is that of Slater and Clark (2006). In their snow data assimilation work, they mapped observations of snow water equivalent from nature into their assimilating model's phase space. First, the observations were converted into percentiles of their long-term climatological distribution, then the snow water equivalent value corresponding to the same percentile within the model's long-term climatological distribution was determined.

Slater and Clark (2006) found that the mapped quantities (percentiles) were spatially more homogeneous than the actual snow observations (which can vary greatly from site to site even at short distances) and made the interpolation within their DA scheme much easier. They cited this ease of interpolation as their primary motivation for using a mapping algorithm. Nevertheless, important elements of the mapping paradigm are utilized in their effort. Note, however, that Slater and Clark (2006) did all their modeling and evaluation work in the framework of the mapped observations, without ever going back to the space of real observations by remapping the forecast values, since they expect that their conceptual model can be better linked with hydrologic stream flow prediction models this way.

In a study on the assimilation of soil moisture, Reichle and Koster (2005), using a procedure similar to that of Slater and Clark (2006), matched the observed and forecast cumulative distribution functions (cdf) to “scale” the observed soil moisture information into the space of the variability of the soil component of a numerical Earth modeling system (percentile matching, see also Reichle and Koster 2004). The “scaled” observations were then successfully assimilated using an Ensemble Kalman Filter algorithm. The scaling algorithm used by Reichle and Koster (2005) is analogous to the mapping paradigm described in this paper. The concept of mapping, however, is applied only in the context of soil moisture data assimilation while using observed or traditionally analyzed data for the rest of the Earth modeling system. Reichle and Koster (2005) note that the moderate success of their data assimilation procedure hinges on many factors and that they could not evaluate the various aspects of their scheme separately.

In a review study presented at a recent scientific meeting, Toth et al. (2005, personal communication) distinguished between two goals for data assimilation: replicating nature as truthfully as possible; and providing states that when used as initial conditions will lead to the best numerical forecasts. They suggested that when assimilating observational data with the aim of initializing imperfect numerical models, one “use only data to the extent it is representative in the modeling system”, and “do not correct for systematic model errors”. These are some of the principles that form the basis of the mapping paradigm described in the present paper.

In a discussion similar to that of Toth et al. (2005), Baek et al. (2006) speculate that “in some cases, it might be desirable to let the forecast model state follow its own attractor”. Some of the approaches they tested in their recent data assimilation study with a simple model differ from the fidelity paradigm in that it allows the state used to initialize background forecasts to deviate from the observations reflecting model related errors. In particular, they remove the estimated forecast bias from the initial condition of the background in their data assimilation cycle; then they add the estimated forecast bias back to the forecast fields prior to performing the data assimilation step.

Baek et al (2006) linearly add various types of artificial biases into their background forecasts that they then attempt to estimate using a Local Ensemble Transform Filter (LETF, see Ott et al. 2004). For estimating the state of the system and the linear bias, they augment the state variables with additional sets of variables that can represent the various forecast biases. Baek et al. (2005) find that with an ensemble size equal to the augmented number of variables in their analysis domain, after a possibly long “settling time” the filter takes to converge, the linear bias can be well estimated.

One of the methods of Baek et al. (2006) shares some important elements with the mapping paradigm described in this paper, i.e., the removal of estimated model related error from forecast initial conditions, and the addition of the same error to the forecast itself following the integration. However, the methods tested in Baek et al. (2006) differ from those introduced in the present study. For example, the methods of Baek et al. (2006) are designed to correct a linear model bias and have not been tested in the general case of nonlinear model-related error behavior that typically emerges in real applications and are studied in the present paper. Also, they estimate model error as part of an ensemble-based data assimilation algorithm, potentially requiring an ensemble size equal to multiples of the size of the state vector, that is impractical for most real life applications, as compared to the easier to apply, dynamically based mapping vector estimation algorithms described in sections 4 and 5 in this study.

3. Numerical modeling and data assimilation tools

3.1 Numerical model

The mapping algorithm will be demonstrated using two versions of the Lorenz (1963) model:

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y-x)\\ \frac{dy}{dt} &= rx-y-xz\\ \frac{dz}{dt} &= xy-bz\end{aligned} \qquad (1)$$

To represent nature, the standard parameter values are used: σ=10, b =8/3, and r =28. To create an imperfect model of nature, two aspects of the model used for nature are modified: (a) z is altered by a constant offset Δz;

$$z_{model} = z + \Delta z, \qquad (2)$$

where the offset Δz = 2.5 (~5% of the standard deviation of z), and (b) by making σ=9. Change (a) shifts the model attractor in the z-direction of the phase space to create a bias in the model, while change (b) slightly modifies the general behavior of the model. Integrations were carried out using the 4th order Runge-Kutta numerical scheme with a time step of 0.01.

3.2 Generation of nature and the forecasts

The Nature representation was generated as a long integration of the original model, started with an arbitrary initial condition close to the attractor. After discarding the initial portion of the trajectory to reduce the influence of model drift, forecasts were initialized every 15th time step. Each experiment (except as noted) consists of 1000 independent forecast cases. Forecast error, e, is defined as the distance (root mean square, or RMS error) between a forecast (x^f) and the corresponding state of nature (x^n) in phase space: $e = \| x^f - x^n \|$.

3.3 Model initialization

Forecasts were started using three different initialization schemes:

- No observational errors: where the model is started with an initial state taken directly from the nature run (Analysis (A) = Observations (y) = Nature (N));
- Replacement method: where observations (y) of all 3 model variables, generated by adding normally distributed random noise (with a zero mean and standard deviation equal to 2) to the nature run, are used as initial conditions (A=y);
- Analysis from a 3DVAR scheme.

The 3DVAR scheme (e.g., Kalnay 2002), given certain assumptions, optimally combines observations (y) and short-range forecast fields (x^f, also called “background”) to produce an analysis (x) that is closer to nature than either the observations or the forecasts themselves. This is achieved by minimizing the cost function *J*:

$$2J = (x - x^f)^T B^{-1}(x - x^f) + (y - H(x))^T R^{-1}(y - H(x)) \qquad (3)$$

where *H*, the observation operator, is a function that maps model values to the phase space of the observations, *B* and *R* are the background and observational error covariance matrices, and superscript T stands for transpose.

Fifteen time steps was the chosen length for the background forecasts in this study. Since initial and model errors with the chosen size reach close to saturation values at around 750 time steps, this choice corresponds to a weather forecast (that typically loses predictability around 15 days) of about 7 hours. B was computed so that:

$$B = \overline{(x^f - x^n)(x^f - x^n)^T} \qquad (4)$$

where x^f is the forecast vector at the15th time step, and x^n is nature valid at the same time, while the overbar denotes time averaging, in this case over 5,000 analysis/forecast cycles from a time period independent of the evaluation period, using the traditional paradigm. Note that B, as defined above, is influenced by both initial value (chaotic) and model related (drift) errors. *B* was empirically tuned by multiplying it with a constant number to decrease the analysis and background errors over the evaluation period. Since observational errors were chosen to be independent with an expected value of 2, the observational error covariance *R* was set to be diagonal with the same value of 2. Note that the simplicity of the system used in this study allows for a better optimization of the 3DVAR method as compared to analysis applications in more complex systems like in Numerical Weather Prediction (NWP, i.e., fewer parameters and larger data samples available for tuning).

When 3DVAR is used according to the mapping paradigm, y in the 3DVAR equation is replaced by M(y):

$$2J = (x - x^f)^T B^{-1}(x - x^f) + (M(y) - H(x))^T R^{-1}(M(y) - H(x)) \qquad (5)$$

where M is a mapping vector that links points on or near the nature attractor to corresponding points on or near the model attractor.

3.4 Estimation and removal of systematic errors

In some experiments with the traditional paradigm, systematic forecast errors will be statistically corrected. The systematic error (S) will be estimated for each forecast lead-time a posteriori, using dependent data, as a difference between the time mean of the forecast and the corresponding analysis fields over the verification period:

$$S = \overline{F - A}\,.$$

Bias-corrected forecasts (F_c) will be generated by removing the systematic error from the forecasts:

$$F_c = F - S \ .$$

The use of independent training data in the estimation of systematic errors would introduce sampling errors, degrading the quality of bias-corrected forecasts. Therefore the results that will be presented with the use of a dependent training dataset will provide an upper bound estimate for the level of error reduction that can be achieved with the use of bias-correction algorithms in realistic applications.

4. Climate mean mapping

The mapping paradigm, as introduced in Section 1, involves

a) Mapping observations close to the model attractor;
b) Data assimilation using the mapped observations;
c) Generation of numerical forecasts from the mapped initial condition;
d) Remapping the analysis and forecast onto the attractor of nature.

The new elements in the mapping paradigm are steps (a) and (d); steps (b) and (c), apart from the changes in inputs, are the same data assimilation and forecast procedures used in conventional studies. Before the new procedure can be applied, however, a challenging task, the estimation of the mapping vector (M) must be accomplished. Once an estimate of M exists, both (a) and (d) are simple operations that require minimal computational resources.

4.1 Estimation procedure

In this section we assume that the long-term climatologies of nature (N) and the model (a free model integration, F) are readily available. The mapping vector is estimated as a difference between the long-term mean of nature and its model:

$$M = \overline{H(F)} - \overline{H(N)}$$

. (6)

In this study, M was estimated based on 247,500 time step integrations of the “nature” and “model” versions of the 3-variable Lorenz system. This would be equivalent to 13-14 years worth of weather data, comparing the error growth characteristics of the Lorenz model to those of atmospheric models. The use of a significantly smaller number of time steps would result in estimates with relatively large sampling errors. When M, as in this study, is to be applied on model variables, the observation operator H is an identity matrix.

4.2 Perfect observations

For the sake of simplicity, the mapping paradigm will be demonstrated first in a situation with perfect observations. Note that in this case all forecast errors originate from the use of an imperfect model. Therefore, the possible positive effect of mitigating model-related errors on the total forecast error is expected to be largest in this scenario. Note that the schematics in Figs. 1 and 2 were numerically generated under this scenario.

The different curves in Fig. 4 show the RMS error of four different forecasts: the continuous and dotted blue lines correspond to the RMS error of a conventional forecast with an imperfect model, started with initial conditions taken directly from nature, before and after bias correction, respectively. The red and green curves show RMS error for forecasts with initial conditions derived by mapping nature using the mapping vector, before and after remapping, respectively.

A few observations follow:

1) Unlike the blue continuous (conventional method without bias correction, starting from nature, no observational error) and the other curves, the red error curve (mapped forecast without remapping) starts at a finite value, reflecting the application of the mapping vector on the initial conditions.
2) Beyond a short initial period, the red curve is below the blue curve for all lead times until the level of nonlinear error saturation is approached. This is because the growth rate of errors for the mapped forecast (red curve) is lower than that for the conventional forecast (blue). This indicates that the application of the mapping algorithm reduces errors introduced from the use of an imperfect model.
3) The green error curve (remapped forecast) is below the red curve (mapped forecast) at all lead times, with the largest differences occurring at short lead times. This indicates that the remapping procedure is successful in moving the forecasts from the attractor of the model back to the vicinity of the attractor of nature. The application of remapping is most important at short lead times, where the displacement caused by the application of the mapping procedure is the largest in a relative sense when compared to the total forecast error.
4) Most importantly, the green error curve (remapped forecast) is below the blue curve (conventional forecast) at all lead times. Forecast error after the first time step, for example, is reduced by a factor of 3. This indicates that the application of the mapping paradigm with the Lorenz model had an overall positive effect on forecast quality.
5) For the first 3 time units the dotted blue curve (bias-corrected conventional forecast) is below the continuous blue curve (raw conventional forecast). The error for the remapped forecasts (green curve), however, is below the

dotted blue curve at all lead times. This indicates that the bias removal, though efficient in improving short-range forecast performance, cannot reproduce the effect of the mapping algorithm. The effect of the mapping algorithm is not due to a simple statistical bias correction, but rather is a result of the elimination of dynamically developing forecast errors triggered by the drift in conventionally initialized forecasts.

The above results can be summarized as follows:

- Steps a and b, the application of the mapping vector to the observations, move the initial condition away from nature (observational point 1);
- Step c, starting the integration of the forecast model from the mapped initial condition, reduces model related errors (point 2);
- Step d, remapping the mapped forecasts, eliminates the initial negative effect of mapping on forecast quality, producing forecasts superior to the conventional forecasts, even after they are bias corrected (points 3, 4, and 5).

To understand the overall positive effect of the mapping algorithm, Fig. 5 displays a segment of nature and a representative series of short (2 time-step) conventional and remapped forecast trajectories. One observes in all cases that the blue (conventional forecast) trajectories display a rapid divergence from nature (black curve), a behavior characteristic of model related errors (Toth and Vannitsem 2002). By contrast, the green (remapped forecast) trajectories closely follow nature, indicating that the mapping algorithm can eliminate a large part of model drift-induced errors and in the absence of observational errors (i.e., initial errors are exclusively due to the error in the estimation of the mapping vector, and consequently are small) produce very high quality short-range forecasts.

The above results can also be interpreted in terms of improved shadowing performance (see, e.g., Grebogi et al. 1990; Judd and Smith 2004). For example, the time limit for shadowing nature with errors as small as the average error in one time step conventional forecasts is extended approximately three-fold.

4.3 Imperfect observations, replacement technique

In this subsection, the effect of the mapping algorithm will be studied in the presence of observational errors, using the same climate mean mapping vector as in Section 4.2. In step (b) of the mapping algorithm the simplest initialization technique will first be used, assuming all three variables of the Lorenz nature model are observed, and initializing the model with the observed values. The results in Fig. 6 are shown in the same fashion as in Fig. 4 (except omitting the line for post-processing). The results in Fig. 6 are consistent with those in Fig. 4 except that:

1) The blue (conventional) and green (remapped) forecast error curves start at a finite value. This is consistent with the presence of observational errors that are also introduced into the initial conditions.
2) The differences between the three curves are much reduced. For example, the remapped forecasts (green) show only about a 15% error reduction (as compared to 67% in Fig. 4) with respect to the blue curve. This is due to the presence of large initial errors that mask the reducing effect of the mapping algorithm on model related errors (which are small when compared to the initial errors). In such a situation, the reduction of model related errors results in relatively small overall forecast improvements.

Notwithstanding the smaller differences between the curves in the presence of initial errors (Fig. 6), the positive effect of the mapping algorithm in reducing the initial growth rate of errors is as large as in the case with no initial errors (Fig. 4), being around a factor of three.

4.4 Imperfect observations, 3DVAR

In this subsection, the effect of the mapping algorithm will be tested with a data assimilation procedure. Unlike in the idealized experiments of Sections 4.2 and 4.3, where the mapping vector was estimated assuming nature is known, nature is replaced in Eq. 6 by analysis fields of nature:

$$M = \overline{H(F)} - \overline{H(A)} \qquad (7)$$

M is estimated using a 75,000 time-step long free model integration and traditionally generated analysis fields associated with a similarly long trajectory of nature (i.e., 5000 analysis fields). Otherwise the same experiments will be carried out as in Section 4.3, except that in step (b) of the mapping algorithm a 3DVAR procedure will be used in place of the replacement technique. The results, shown in Fig. 7, are consistent with those in Fig. 6 except that:

1) The initial error for all methods is lower than when the replacement technique is used due to the beneficial noise reduction properties of the 3DVAR DA technique;
2) More importantly, the starting point for the green curve is 9% lower than that for the blue curve. This shows that the remapped initial condition (analysis) is of higher quality than the conventional analysis.

Point (2) above indicates that the combination of higher quality first guess fields from the mapping algorithm and the mapped observations leads to analysis fields that when remapped back to nature better capture nature than traditional analysis fields generated based on the fidelity paradigm.

5. Adaptive mapping

For some dynamical systems the mapping vector may exhibit either regime and/or seasonally dependent variations; or for other applications no long-term climatological mean for nature or the model may be available. In such cases, using the mapping vector estimation procedure based on the long-term climatological mean difference between nature and the model, presented in the previous section, would yield sub-optimal or poor results. For such applications, an adaptive mapping vector estimation algorithm is presented and tested in this section.

5.1 Estimation procedure

The adaptive mapping vector estimation is an iterative procedure that requires only a small amount of training data to begin with. In the course of the iterations, the initial crude estimate of the mapping vector is gradually refined. Once the analysis fields have asymptotically converged to the model attractor, further variations in the mapping vector estimate reflect the regime-dependent and/or cyclic behavior of the mapping vector that can potentially yield better analysis/forecast results when compared to a climatologically fixed mapping vector estimate.

The adaptive mapping vector estimation procedure is part of the mapping algorithm. The algorithm is started by running conventional data assimilation/forecast cycles based on the fidelity paradigm, i.e., M = 0. After a relatively small number of analysis/forecast cycles the first increment of the mapping vector, M_{Incr}, is computed as the time mean difference between the analysis (A) and corresponding first guess (F_g) fields:

$$M_{incr} = \overline{H(F_g)} - \overline{H(A)} \qquad (8)$$

This difference vector reflects the model drift that occurs during the length of a forecast integration within the assimilation cycle. For the next iteration period, the mapping vector estimate used in the previous iteration (M_{prior}) is updated by M_{Incr} such as:

$$M = M_{prior} + M_{Incr} \quad (9)$$

M is used as the mapping vector in the mapping algorithm during the second iteration period, moving the observational data closer to the model attractor and using its reverse to move the analysis and forecast fields back toward nature for applications, including verification/evaluation. Since the first iteration(s) of M is(are) not expected to bring the observations all the way to the model attractor, subsequent iterations of the procedure are needed, and at the end of each M_{Incr} is determined and M is updated as in Eqs. (8) and (9). This iterative process will let M approach its asymptotic behavior. In cases where the mapping vector is constant on the attractor, M is expected to converge to that value. Otherwise, the estimate of M will fluctuate, reflecting the regime dependent or cyclic behavior of the mapping vector.

The adaptive mapping vector estimation method described above consists of the following steps:

a) Set $M_{prior} = M = 0$;

b) Use M in mapping algorithm during next iteration period;

c) Based on a small sample collected during the iteration period, determine

$$M_{incr} = \overline{H(F_g)} - \overline{H(A)}$$

d) Update M by $M = M_{prior} + M_{Incr}$.

Steps b-d are repeated for each iteration of the method.

The rate at which M converges to its asymptotic behavior depends on the characteristics of model drift; in particular, how much of the drift occurs during the length of a forecast used in the data assimilation cycle (i.e., 15 time steps in this study). The length of the iteration period (i.e., the number of assimilation/forecast cycles) must be made long enough to effectively filter out noise in the computation of M_{Incr}. Once the estimate of M converges, variations in M can be monitored and used on a continual basis to track regime dependent changes in the mapping vector.

5.2 3DVAR application

In this subsection, the adaptive mapping vector estimation will be tested with a 3DVAR scheme assimilating imperfect observations. The experimental design is similar to that used in Section 4.4, except for the use of the adaptive mapping vector estimation procedure in place of climate mean mapping. For each iteration, analysis and first guess data from 120 analysis/forecast cycles, with a cycle length of 15 model time steps, will be used. In terms of weather forecast applications the length of the iteration period is around 34 days, which is comparable in length to the periods used for estimating regime dependent systematic errors for bias correction of short-range forecasts (e.g., Stensrud et al., 2003; Cui et al., 2006).

To demonstrate the convergence of the mapping vector estimate based on the adaptive procedure to its asymptotic behavior, Fig. 8 shows the three components of the mapping vector estimate M for the first 16 iteration periods. During the first few iterations, the series of mapping vector estimates for variable z display a strong convergence to a range close to the climate mean mapping vector estimate. The convergence, corresponding to the drift of the analysis toward the model attractor, appears to be complete by around the 5th iteration (which would correspond to around half year's worth of data when compared to weather forecasts). The drift is less apparent for the x and y variables which fluctuate around values slightly above or near the climate mean mapping vector estimate. Note the relatively low frequency variations observed in mapping vector estimates, especially for the x and y components of the system, that must reflect nonlinear interactions.

The relatively quick rate of convergence is also evident when the quality of the analyses and forecasts from the mapping algorithm is monitored through the iteration steps. Starting from the third iteration step (using a mapping vector estimate based on only 120 cases, or a month's worth of data for weather forecasts), the skill of the remapped analysis and forecast fields shows a clear advantage over those from a conventional analysis/forecast cycle (not shown).

The skill of the remapped analyses and forecasts beyond the convergence period is displayed in Fig. 9. The results confirm that the adaptive method provides practical and useful estimates of the mapping vector.

One can ask whether the mapping algorithm for the Lorenz model is more skillful with the climatologically fixed vector or with an adaptive estimate of the mapping vector. Fig. 10 compares the errors in the remapped forecasts with the climate mean versus the adaptive mapping vector estimates for the same set of cases. Interestingly enough, the adaptive method yields more skillful remapped analyses and forecasts. The differences in the quality of analyses and short-range forecasts are statistically highly significant. Assuming that the analyses (separated by 15 time steps) and those ensuing forecasts that are non-overlapping (e.g., every third of the 3 time unit lead-time forecasts on Fig. 10) are independent, the differences are statistically significant at the analysis and up to 2/10 time unit lead-time forecasts at the 0.01/10% or higher level. These results indicate that:

1) The mapping vector for the Lorenz model is not constant but rather varies over the attractor; and
2) The adaptive method presented in Section 5.1 is capable of capturing at least some of these regime-dependent fluctuations, leading to improved analysis/forecast performance.

6. Summary

This study addressed how to best choose initial conditions when numerical forecast integrations of nature are made with imperfect models. In an attempt to reduce initial value related errors, forecasts have traditionally been initialized with states as close to nature as possible (a “fidelity” paradigm). An initial state close to nature, however, will necessarily be off the attractor of an imperfect model, leading to model drift and associated forecast errors. To reduce drift-induced errors, a new concept called the “mapping” paradigm was proposed in place of the fidelity paradigm.

The new paradigm is based on the assumption that a one-to-one mapping exists between points on or near the attractor of nature and corresponding points on or near the model attractor. According to the new paradigm, observations are moved by a mapping vector from the vicinity of the attractor of nature to that of the model. Data assimilation is performed using the mapped observations and the resulting “mapped” analysis is used to initialize numerical forecasts that are also used as first guess fields in the next data assimilation cycle. For general applications, the analysis and forecasts are remapped back to the vicinity of the attractor of nature with a vector opposite from the mapping vector. The expectation is that performing a numerical integration near the model attractor will significantly reduce the error that would otherwise arise due to a drift of model integration trajectories started from a point near the attractor of nature to that of the model.

Two algorithms were proposed for estimating the mapping vector. The conceptually simple climate mean estimation method defines the mapping vector as the difference between the long-term climatological means of nature and the model. If the mapping vector is expected to exhibit regime-dependent or cyclic behavior, or no long series of states are available for either nature or the model, an adaptive method is proposed that requires an order of magnitude less data.

The new mapping paradigm was tested using a Lorenz (1963) 3-variable dynamical system as nature and a modified version of it as the imperfect model. Experiments were carried out in the absence and presence of observational errors, with a simple observation replacement and a 3DVAR data assimilation method as initialization techniques. The main results of this study are as follows:

- In the absence of observational uncertainty, the mapped forecasts initialized from a state intentionally moved away from the attractor of nature by the climate mean mapping vector and then remapped with a vector opposite from the mapping vector to the vicinity of nature exhibited an error growth slower by a factor of three compared to a traditional forecast. A statistical bias correction of the traditional forecasts had only a limited effect as compared to the performance of the remapped forecasts.

This result proves the concept of the mapping paradigm. It shows that in case of the Lorenz model, mapping exists and with a simple procedure it can be successfully estimated and used to reduce drift-induced forecast errors in practice.

- When 3DVAR analysis fields generated with a data assimilation / forecast cycle using the mapping algorithm are remapped back to the vicinity of nature, they show significantly less error with respect to nature compared to analyses from a traditional data assimilation cycle based on the fidelity paradigm.

This result indicates that the ability of the mapping algorithm to reduce drift-induced errors is useful in reducing not only forecast but also analysis errors. Paradoxically, moving the observational data away from nature and creating the numerical forecast used in the data assimilation cycle near the attractor of the model (and not near nature as done traditionally) improves the fidelity of the remapped analysis fields.

- The mapping algorithm with an adaptive mapping vector estimation produced higher quality analysis and forecast fields compared to those made with the climate mean mapping vector estimation method.

This result indicates that the mapping vector in the Lorenz system varies by phase space location and such variations can be successfully monitored with an adaptive estimation method using a relatively small data sample.

7. Discussion

7.1 Justification of the concept

To some readers the concept behind the mapping paradigm i.e., moving the initial conditions away from observations, and the validation results using the Lorenz 3-variable system may at first sight appear counterintuitive or unexpected. If initial and model errors are not separated at least conceptually, one may be tempted to follow the traditional paradigm and attempt to force the analysis to stay close to nature despite the natural tendency of first guess forecasts to drift away from nature toward the model attractor.

The mapping paradigm is, however, consistent with general considerations of dynamical systems theory. To realize its potential merits, one must first recognize that choosing an initial state close to nature which is clearly off of the model attractor (the fidelity paradigm), will lead to the emergence of model drift and related forecast errors and, therefore, is not the best way to initialize an imperfect model. Second, if forecasts initialized even in such a sub-optimal way as the traditional paradigm (i.e., the initial condition is off the model's attractor) turn out to be useful, that in itself is an indication of the existence of a mapping between nature and the model. Third, one does not need a perfect estimate of the mapping vector. The quality of traditional analyses and forecasts can be surpassed as long as the mapping algorithm brings the initial condition and forecast *closer* to the model attractor than they would be by using traditional initialization procedures (see Fig. 8 and associated discussion). This is true especially since the analysis and forecast fields are remapped back toward nature using the opposite of the mapping vector.

It is the apparent complexity of the problem that may have deterred some investigators from a full exploration of the optimal choice of initial conditions for imperfect models. The mapping paradigm is based on the recognition of the differences between initial and model related errors. Once the different origins and behaviors of initial condition and model related errors are clarified, it becomes evident that to reduce model drift-induced errors, one must search for initial states that represent nature on or near the model attractor. This is a simple and probably old idea, with some related limited experiments reported in the literature. The real challenge is to find a general procedure for identifying the mapping vector that can be used in practice to reduce drift-induced errors and thus improve forecast performance.

7.2 Mapping relationship

In this paper, two approaches for estimating the mapping relationship between states on or near the attractor of nature and the model were proposed and tested, but other approaches may be designed as well. Some readers may question how it is possible to arrive at useful estimates of the mapping vector by using the adaptive method with relatively small amounts of data (based on 120 cases only), while much larger data sets are necessary for estimates of similar quality based on the climate mean difference method.

One must note that when the climate means of nature and the model are computed, there is no one-to-one relationship between the states considered from the two sources (i.e., long segments from nature and a free model run). Therefore, any systematic difference can be evaluated only in a statistical sense, by taking the mean of a large number of realizations, preferably covering a large part of the attractors. This is necessary to avoid the large sampling errors that may result if different parts of the nature and model attractors are not represented in the same proportions in small data samples.

By contrast, the adaptive method compares analysis and short-range forecast states valid at the same time. Unlike the randomly chosen trajectory segments that are compared in the climate mean procedure, these states are designed to closely shadow nature and hence the analysis – forecast pairs valid at the same time are directly related. Their difference field is therefore dominated by systematic and not by chaotic type influences. This allows for an incremental approximation of the full mapping vector, starting with analysis fields close to nature (the traditional analysis) that, after a number of iteration periods, approach the model attractor, without ever losing the shadowing relationship between nature and the analysis fields.

Once a good estimate of the mapping vector is attained, the mapping algorithm can increase the time limit for which shadowing is possible with an imperfect model. While the use of mapping leads to a rather significant 3-fold increase in shadowing time for the 3-variable Lorenz system, the quantitative effect in other systems depending on the Lyapunov spectrum and other characteristics of the systems in question, may be less dramatic. The increase in shadowing time is achieved by allowing the model forecasts to systematically differ from nature as determined by the mapping vector. The smaller the relative role of initial value related errors, the more positive impact the mapping algorithm would have for shadowing and, in general, for reducing total forecast errors.

Whether there exists a unique optimum mapping relationship between particular numerical models and nature is an open question. This question, however, may be irrelevant for practical applications, as long as estimates of mapping provide useful results. As one considers progressively poorer models of nature, the identification of a mapping relationship is expected to become more difficult and beyond a certain point of dissimilarity (at which a model may become completely useless for forecasting) mapping may not exist any more.

7.3 Limitations of the method

As for the limitations inherent in the method, it must be pointed out that the mapping algorithm is designed to reduce forecast errors associated with the *transient process of model drift* that a traditionally initialized forecast experiences. Other model related forecast errors that are independent of any transient behavior (e.g., errors in the speed or configuration of phenomena in nature arising from erroneous/simplified model formulation), referred to as “asymptotic”, are not affected by, but easier to identify with the mapping algorithm.

Another limitation of the mapping algorithm is that to the extent that the estimate of the mapping vector has some error in it, the mapped initial conditions will be closer to but not exactly on a model trajectory corresponding to nature. Depending on the quality of the mapping vector estimation, model drift may be significantly reduced but not necessarily eliminated.

The mapping algorithm presented here attempts to assess the systematic differences between nature and its representation in a numerical model. While it is expected to work with biased observations, better performance in terms of mapping vector estimation and forecast quality is expected if the observations are bias corrected before their use. Techniques for such observational system-dependent bias correction can be readily found in the literature (see, e.g., Derber and Wu 1998).

7.4 Practical applications

The concept of mapping is general. The results presented in this paper provide a proof of concept in the case of a simple system. The algorithms proposed, using either the climate mean estimation or the adaptive estimation (if needed for the description of regime-dependent or cyclic variations in the mapping vector, or because of a lack of climatological data) can be easily applied to more complex systems. For example, the generation of a 13-14 year integration with a model used in operational Numerical Weather Prediction (NWP) for climate mean mapping vector estimation, though not trivial, is a doable task. Adaptive mapping vector estimation methods would need significantly less resources. Whether mapping is applicable and practically useful for complex systems can be answered only through further studies.

From a practical application point of view it is encouraging that systematic differences between analyses and short-range forecasts can be well estimated in real world applications (see, e.g., Dee and Da Silva, 1998; Cui et al., 2006; Glahn and Lowry, 1972). This is because the time-mean difference between analyses and corresponding short-range forecasts is dominated by their systematic difference, not by initial value-related errors that amplify rapidly due to chaotic error growth at later lead times. The adaptive mapping algorithm, in fact, builds on such estimates of systematic differences between analysis and first guess fields. Because only a small portion of the model drift occurs during the length of a single first guess forecast, the adaptive method allows the analysis to “drift” closer to the model attractor with each successive iteration. This is achieved by moving the observations in each iteration with a mapping vector estimate incremented by the systematic difference between the analysis and first guess fields determined over the previous iteration period (see Section 5.1).

As described in the Introduction, estimation of systematic differences between forecasts and a proxy for truth is an important element of both the traditional and the new mapping paradigms. There are two critical differences though regarding the estimation and use of these systematic differences. In the traditional approach, the systematic error is estimated as a function of lead-time and it is removed from the forecasts, to bring the model forecast closer to the attractor of nature. In contrast, the mapping approach estimates an asymptotic, lead-time independent systematic difference, and uses that to move the observations closer to the model attractor.

A possible test bed for the application of the mapping paradigm in more complex systems may be coupled ocean-atmosphere model forecasts of Earth's climate system. First, model related errors for general circulation models of the ocean are generally estimated to be larger than those for the atmosphere. And second, when coupled ocean-atmosphere models are started with observed initial conditions, typically a large drift develops. This is partly due to the inability of the imperfect model components to realistically represent the delicate balances that exist in nature with respect to the heat and momentum exchange between the two sub-systems. The mapping of observational data from both sub-systems could possibly lead to a reduction in model drift and drift-induced forecast errors.

7.5 Potential benefits

If proven useful for complex systems, application of the mapping paradigm may have a number of potential advantages in the areas of weather, climate, or other natural systems:

- Improved forecasts due to the elimination of drift induced errors;

Traditional forecasts start from a state close to that observed in nature. If we adopt the mapping paradigm, such a forecast would be considered to be started with an initial error opposite to the mapping vector. This error may initially decay as the forecast drifts closer to the model attractor. However, the forecast trajectory in general will remain different from the trajectory started from a mapped initial condition. It is this difference, or drift-induced error, that can be reduced or eliminated via the mapping algorithm. When using the mapping paradigm, in the absence of drift-induced problems, ensemble forecast performance would also be improved since ensembles are designed to represent initial value and possibly asymptotic model error related forecast uncertainties but not drift-induced errors.

- Improved analysis fields due to the use of more accurate forecasts in DA;

In the framework of the fidelity paradigm, observational data are used directly as they are, forcing the initial states to be close to nature. Systematic differences between nature and first guess background fields are assessed to move the background even closer to nature (and hence further away from the model attractor). This leads, as described above, to the emergence of additional, drift-induced forecast errors. These errors naturally appear in the analysis fields that use the forecasts as first guess fields. The adaptive mapping algorithm also uses estimates of systematic differences between analyses and forecasts. These systematic differences, however, (1) are assessed not in a lead-time dependent manner, for a particular lead time of the first guess forecast, but asymptotically; and (2) the estimates are used to move the analysis fields not toward nature, but the attractor of the model. With these features the mapping algorithm eliminates drift-induced forecast errors and after remapping, produces analysis fields that are of higher quality than those from the fidelity paradigm.

Though in this study only the 3DVAR method was tested, other DA techniques such as 4DVAR or ensemble-based methods should also benefit, without any special changes, from the use of the mapping algorithm. In fact, ensemble-based data assimilation methods (see, e.g., Tippet et al. 2003) used with the mapping algorithm may be most effective in reducing drift-induced errors by searching for initial conditions that are both very close to the model attractor (via using ensemble-based background error covariance information), and correspond well with the state of nature (via the use of the mapping/remapping procedure).

- Assessment of asymptotic model errors without the masking effect of model drift;

Forecasts with the fidelity paradigm are affected both by the drift of the model and the errors due to initial condition uncertainty. The assessment of asymptotic model errors is severely hindered, and possible only through the *statistical* comparison of a very large sample of unrelated states from nature and a free model run. By contrast, by untangling initial condition and model related errors the mapping algorithm establishes a *one-to-one connection* between states of nature and the model. In the absence of model drift, asymptotic (i.e., lead-time independent) model errors can be readily identified. The estimate of the mapping vector, in fact, can serve as a quantitative measure of asymptotic model errors, that, with the adaptive estimation method can be assessed as a function of phase space location.

The adaptive mapping algorithm can also be used to determine, over a relatively short period of time and without having to use forecast information beyond the length of the first guess, whether proposed changes to a model bring it closer to or move it further away from nature. Tuning may become easier, including empirical or stochastic improvements to dynamical models proposed by, e.g., D'Andrea and Vautard (1999). By making the behavior of the model independent of forecast lead-time, an artificial chasm between weather (dominated by drift related model errors) and climate (dominated by asymptotic model errors) forecasting can also be bridged.

- Eliminating the need for lead-time dependent statistical bias correction.

Drift related errors in forecasts initialized with the traditional fidelity paradigm make the use of lead-time dependent bias correction algorithms imperative. Such schemes, especially for extended lead times, require the costly generation of a long series of analyses and forecasts based on historical data (e. g., Hamill et al., 2003). To compound the problem, in order to assess systematic errors for new versions of the analysis or model, the generation of large reanalysis and reforecast databases need to be repeated after any improvements are made to the system. By contrast, mapping the initial conditions significantly reduces or eliminates model drift, and remapping the forecasts corrects for asymptotic model errors. Therefore the use of the mapping algorithm much reduces or eliminates the need for the logistically complicated lead-time dependent correction of systematic errors.

8. Acknowledgements

Discussions with Ed Epstein in the early 1990s, who challenged the first author to find a way to estimate the mapping vector, and more recently with Guocheng Yuan and other colleagues are acknowledged. Comments by David Parrish, Augustin Vintzileos, Dusanka Zupanski, and Milija Zupanski on an earlier version of this manuscript are greatly appreciated. The manuscript was language edited by Mary Hart. The second author was partially supported by a grant from the Office of Global Programs (OGP) of NOAA.

9. References

Baek, S., B. Hunt, E. Kalnay, E. Ott, and I. Szunyogh, 2006: Local ensemble Kalman filtering in the presence of model bias. Tellus A 58 (3), 293-306.

Cui, B., Z. Toth, Y. Zhu, D. Hou, D. Unger, and S. Beauregard, 2006: The trade-off in bias correction between using the latest analysis/modeling system with a short, vs. an older system with a long archive. Proceedings of the First THORPEX International Science Symposium. December 6-10, 2004, Montréal, Canada, 281-284.

D'Andrea, F., and R. Vautard, 1999: Reducing systematic errors by empirically correcting model errors. Tellus, 52A, 21-41.

Dee, D., and A. Da Silva 1998: Data assimilation in the presence of forecast bias. Quart. J. Roy. Meteor. Soc., 124, 269-295.

Dee, D., and R. Todling 2000: Data assimilation in the presence of forecast bias: The GEOS Moisture Analysis. Mon. Wea. Rev., 128, 3268-3282.

Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Wea. Rev., 126, 2287–2299.

Glahn and Lowry, 1972: The use of model output statistics in objective weather forecasting. J. Appl. Meteor. 11, 1203-1211.

Grebogi, C., S. M. Hammel, J. A. Yorke, and T. Sauer, 1990: Shadowing of physical trajectories in chaotic dynamics: Containment and refinement. Physical Review Letters 65, 1527–1530.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2003: Ensemble re-forecasting: improving medium-range forecast skill using retrospective forecasts. Mon. Wea. Rev., 132, 1434-1447.

Judd, K., and L. A. Smith, 2004: Indistinguishable states II: The imperfect model scenario. Physica D, 196, 224–242.

Kalnay, E., 2002: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, 341 pp.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. Quart. J. R. Met. Soc., 112, 1177-1194.

Lorenz, E. N., 1963: Deterministic non-periodic flow. J. Atmos. Sci., 20, 130-141.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505–513.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zemin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. York, 2004: A local ensemble Kalman filter for atmospheric data assimilation. Tellus A 56 (5), 415-428.

Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. Geophysical Research Letters, 31, L19501.

Reichle, R. H., and R. D. Koster, 2005: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. Geophysical Research Letters, 32, L02404.

Schneider, E. K., B. Huang, Z. Zhu, D. G. DeWitt, J. L. Kinter III, B. Kirtman, and J. Shukla, 1999: Ocean data assimilation, initialization and predictions of ENSO with coupled GCM. Mon. Wea. Rev. 127,1187-1207.

Slater, A.G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. Journal of Hydrometeorology, 7, 478–493.

Stensrud, J. D., and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England. Mon. Wea. Rev., 131, 2510–2524.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square-root filters. Mon. Wea. Rev., 131, 1485–1490.

Toth, Z., and S. Vannitsem, 2002: Model errors and ensemble forecasting. Proceedings of the 8th ECMWF Workshop on Meteorological Operational Systems. November 12-16, 2001, Reading, England, 146-154.

Zupanski D. and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. Mon. Wea. Rev., 134, 1337-1354.

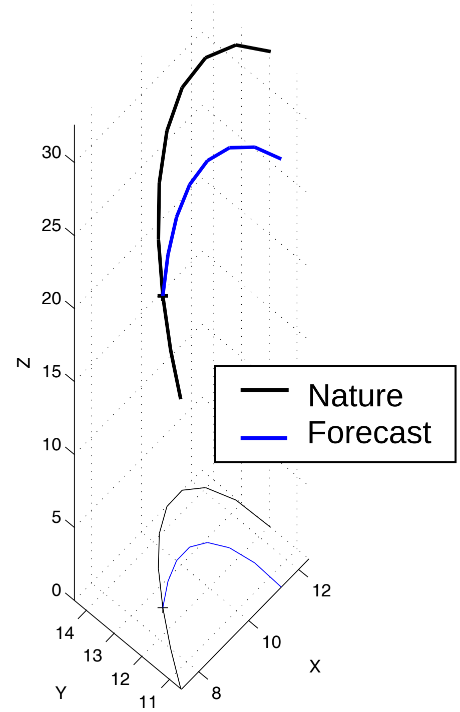

Figure 1. Schematic illustrating model drift in a 3-dimensional phase space, as well as on a 2-dimensional plane of x and y. The black curve represents nature; the blue curve, starting from a point on the trajectory of nature, represents a forecast with an imperfect model. For further details, see text.

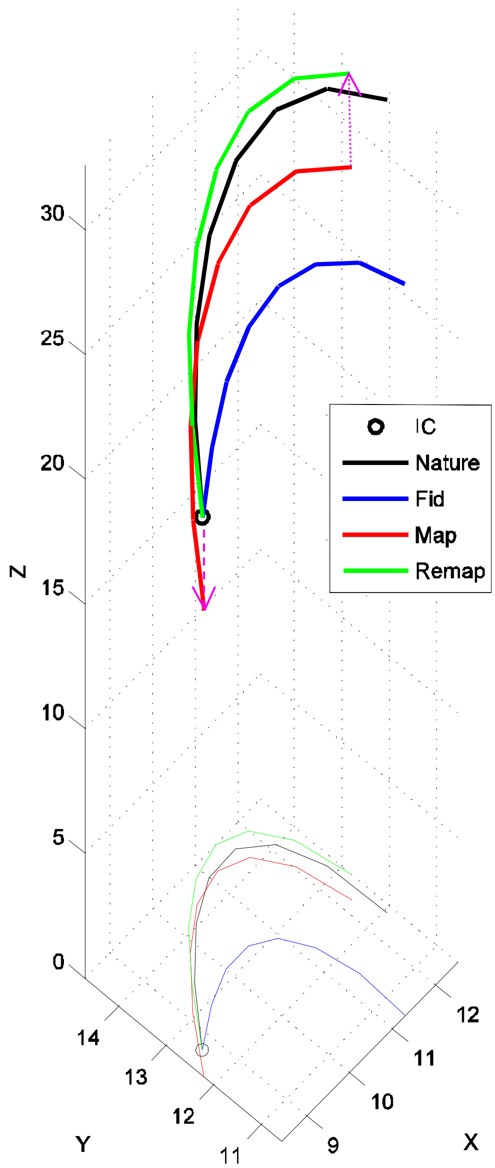

Figure 2. Same as Fig. 1, except illustrating the mapping paradigm, assuming perfect knowledge about the initial state of nature. The black and blue curves are nature and its forecast with an imperfect model; the dashed purple arrow indicates the mapping vector by which the initial state of nature is moved to the model attractor; the red curve represents the mapped forecast; the dotted purple arrow indicates the remapping of the forecast to/near the attractor of nature; and the green curve is the remapped forecast.

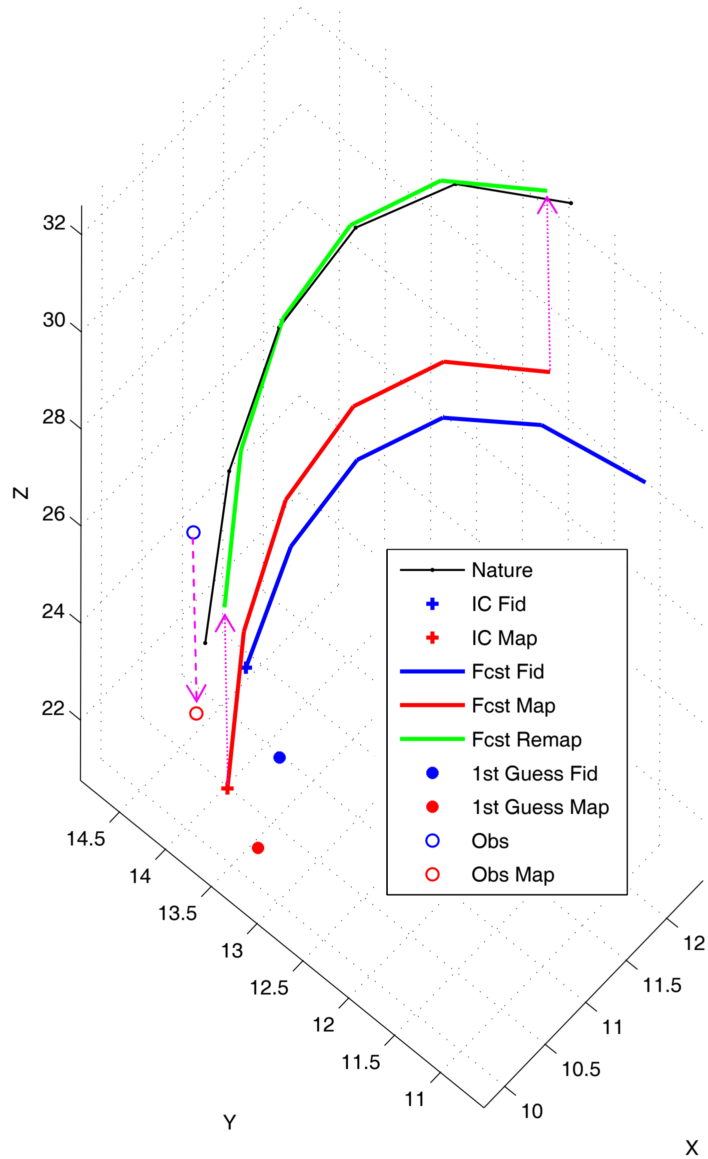

Figure 3. Same as Fig. 2, except with observational errors and data assimilation. The black curve represents nature, with the blue open and closed circles, cross, and curve standing for observations, the first guess, the analysis and forecast from a traditional 3-DVAR analysis/forecast cycle, respectively. The dashed purple arrow represents the mapping vector moving the observations near the attractor of the model, with the red open and closed circles, cross, and curve standing for the mapped observations, the first guess, and the forecast from an analysis/forecast cycle with mapping, respectively. The green curve is the forecast remapped with the opposite of the mapping vector (dotted purple arrow).

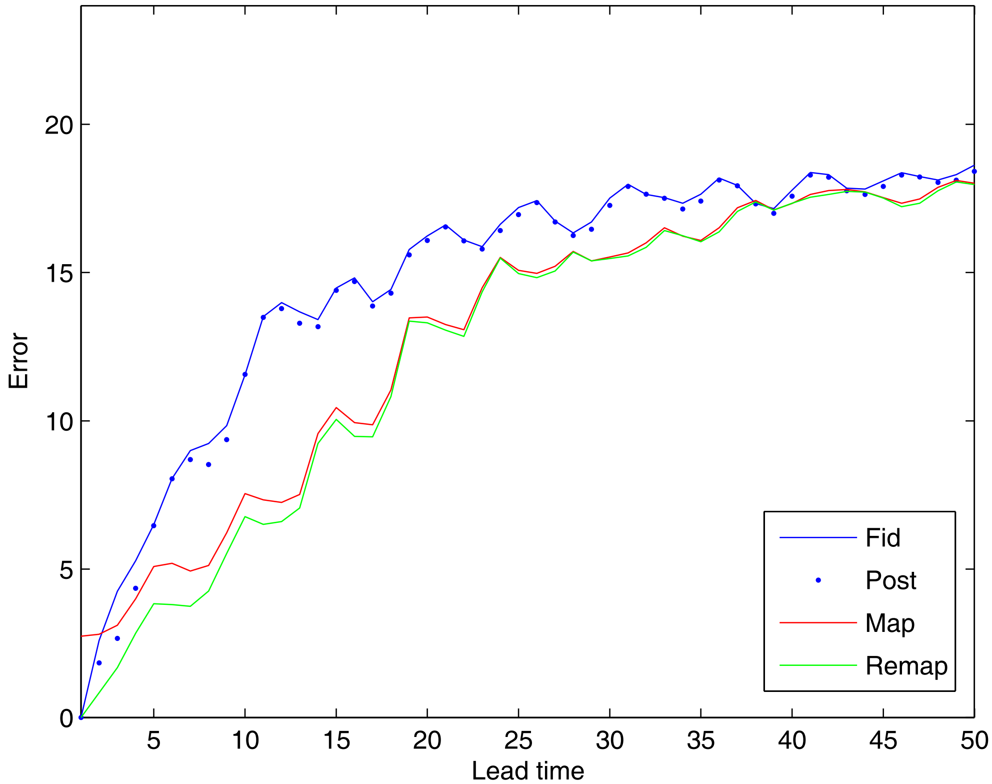

Figure 4. Rms error with perfect initial conditions using the traditional forecast approach before (solid blue) and after statistical post-processing (dotted blue), the mapping algorithm with climate mean mapping vector estimation before (red) and after (green) remapping the forecasts. Time unit is 15 model time steps, and 1 on the horizontal axis corresponds to initial time.

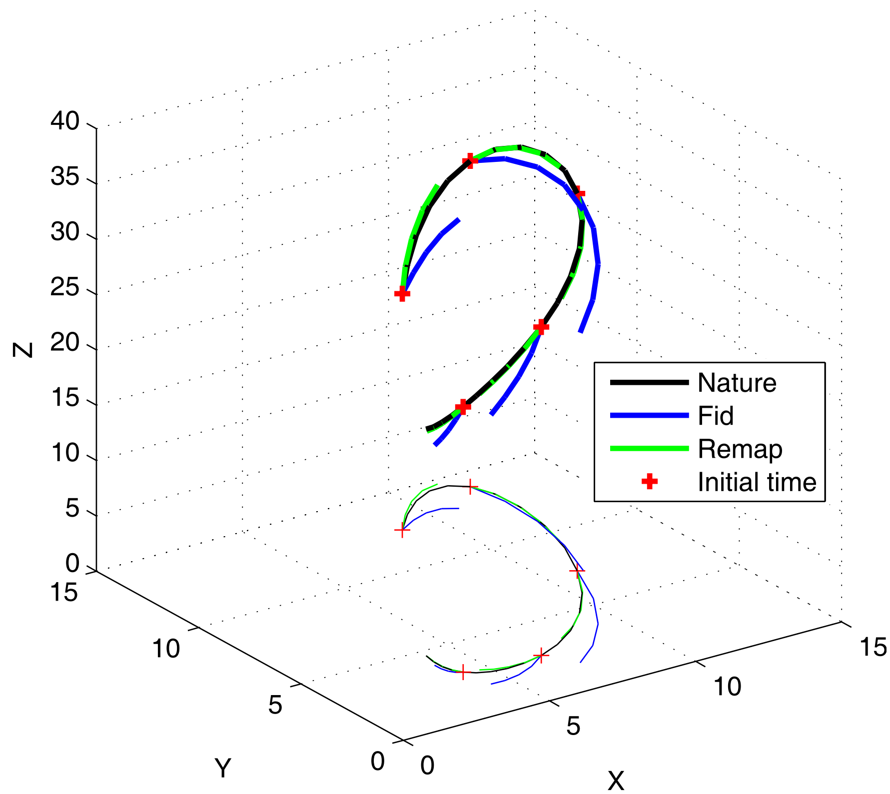

Figure 5. Same as Fig. 2, except for five consecutive forecast segments, indicating the difference in shadowing performance between the traditional (blue) and remapped (green) forecasts with an imperfect model but no initial errors.

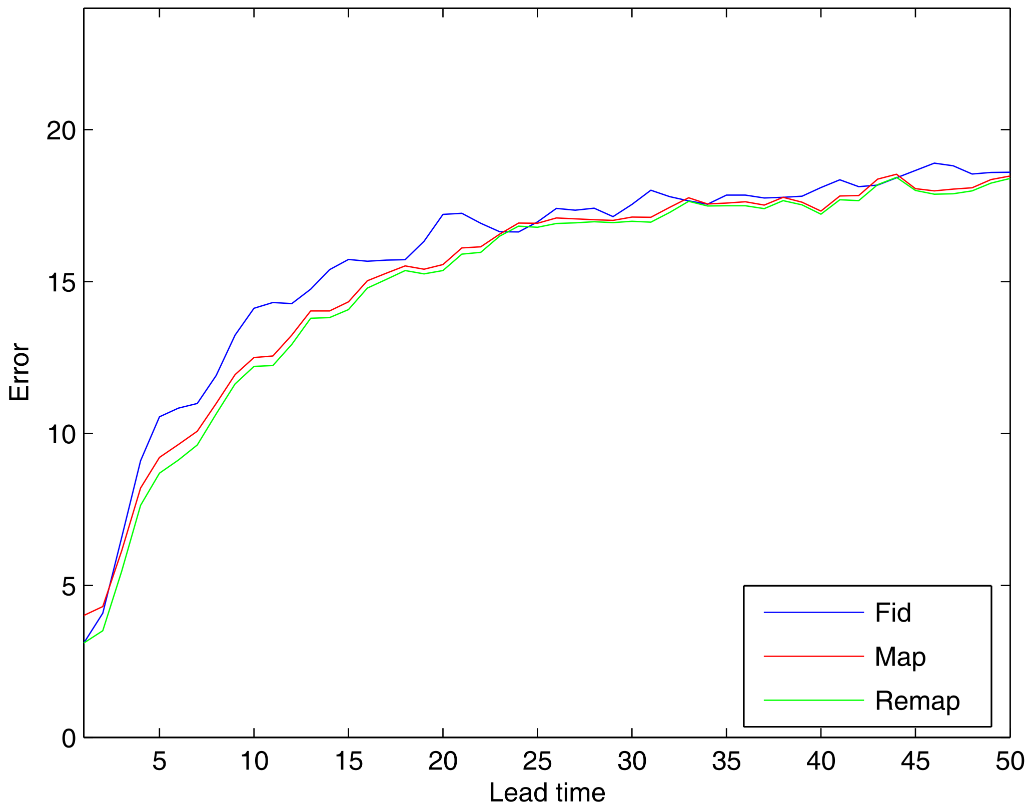

Figure 6. Same as Fig. 4, except in the presence of observational errors, with replacement initialization.

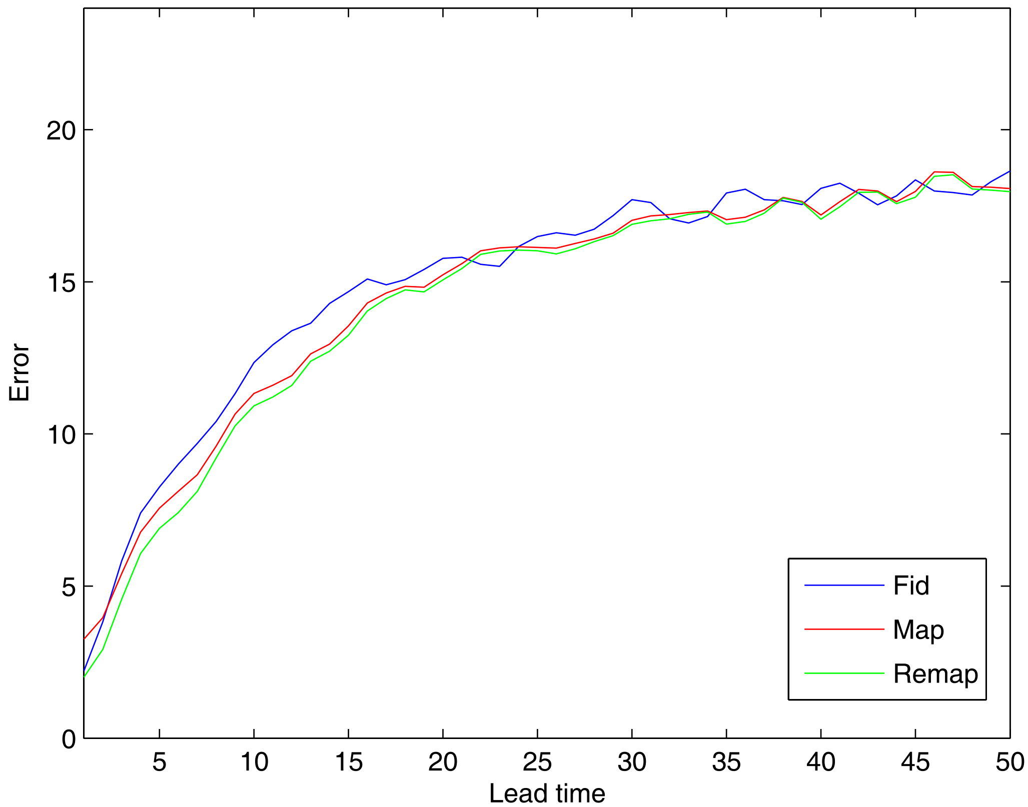

Figure 7. Same as figure 6, except with 3-DVAR initialization.

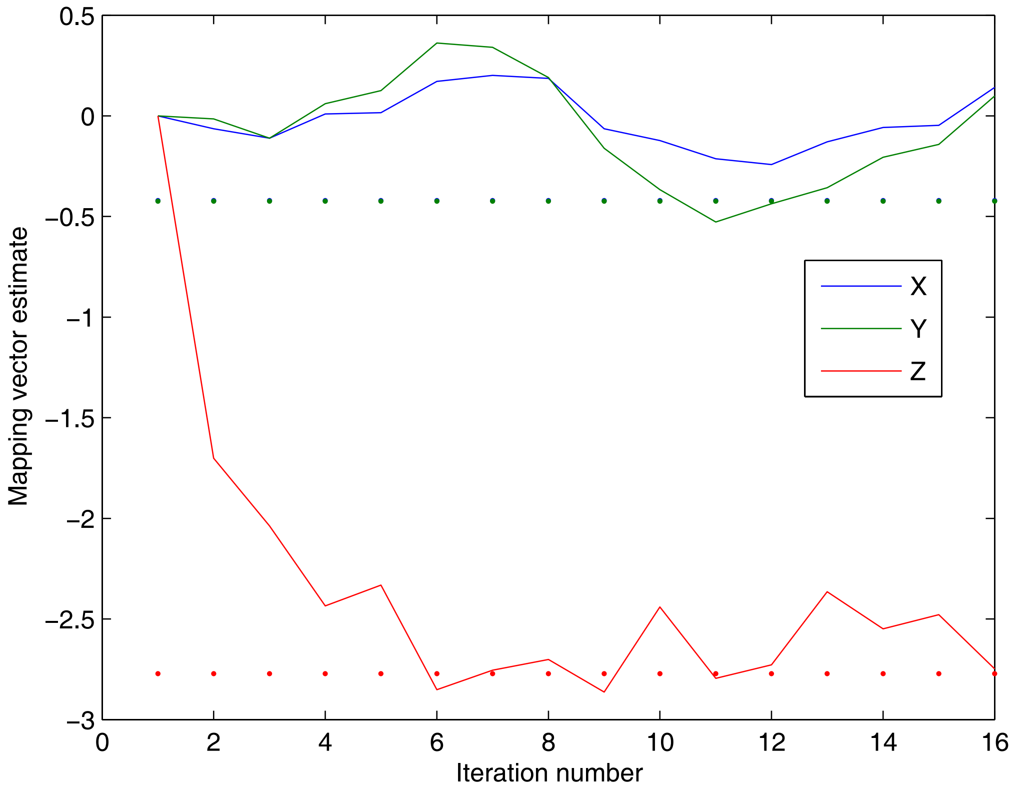

Figure 8. Three (x, y, z) components of the adaptive estimate of the mapping vector M (vertical axis) for iteration periods 1-16 (horizontal axis). The dotted horizontal lines (the green and blue curves are nearly overlapping) represent the time invariant climate mean mapping vector estimate (see Section 4.1).

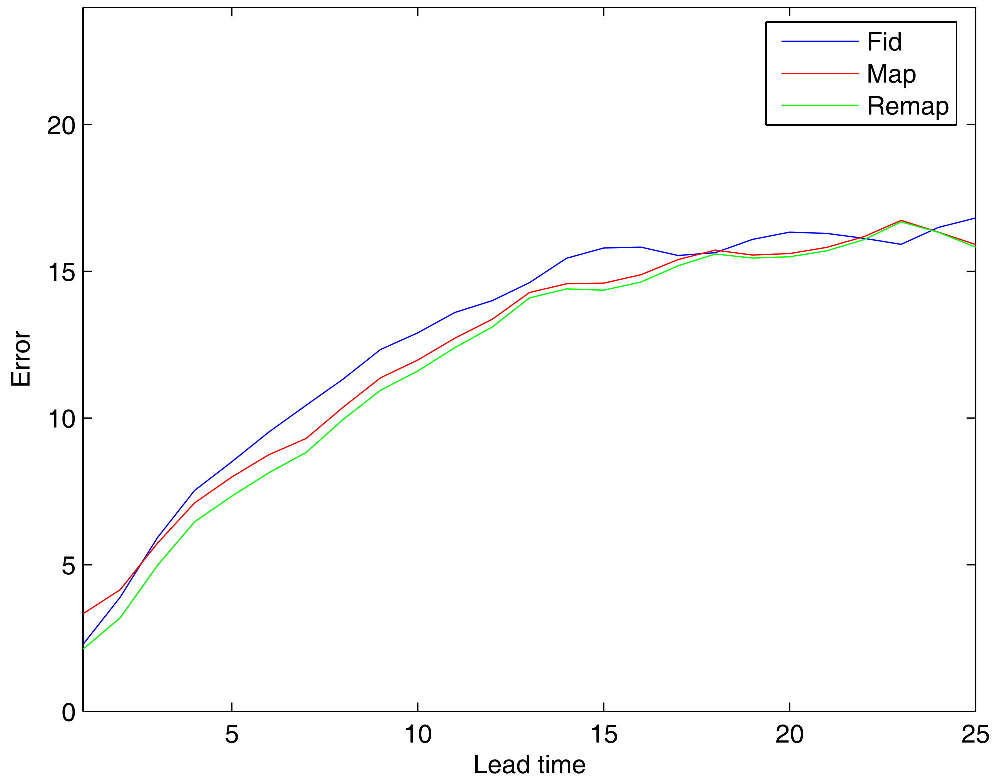

Figure 9. Same as Fig. 7, except with adaptive mapping vector estimation, and averaged for iteration periods 6-16, shown for 25 time units.

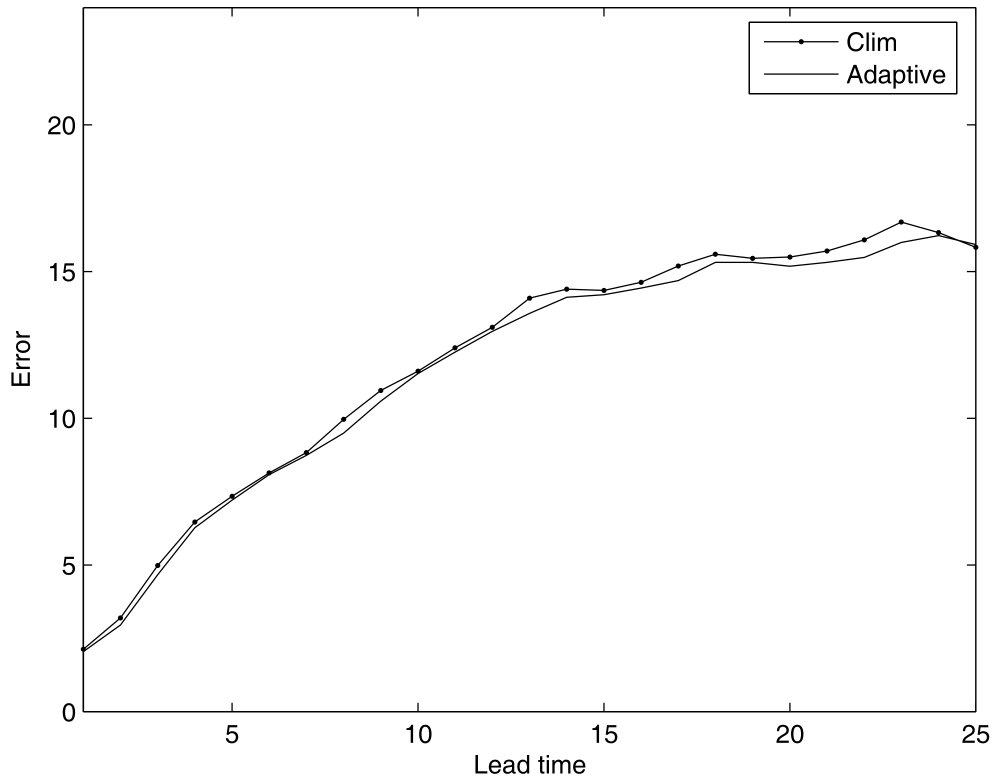

Figure 10. A comparison of errors for the remapped forecasts with adaptive (solid green line, as in Fig. 9) and with climate mean mapping vector estimation (green line with dots, as in Fig. 7), valid for the same cases as in Fig. 9.