Q. J. R. Meteorol. Soc. (2003), **129**, pp. 1255–1267 doi: 10.1256/qj.01.75

Forecast 4D-Var: Exploiting Model Output Statistics

By JAMES A. HANSEN* and KERRY A. EMANUEL
Massachusetts Institute of Technology, USA

(Received 9 May 2001; revised 17 July 2002)

SUMMARY

Forecasts of certain weather elements are improved by linearly relating observed elements to past observations, climatological information, and numerical weather-prediction-model output. Model Output Statistics (MOS) is a statistical post-processing of model output that is capable of forecasting some subgrid scale, synoptically-forced events and of correcting some systematic, but state dependent, model bias. This tendency of accounting for model error is exploited by feeding MOS forecasts back into the state estimation problem. MOS forecasts and their associated uncertainty are treated as 'observations' of the future system state, and a four-dimensional variational assimilation procedure is employed to improve the original analysis and resulting model forecast. In a simple-model scenario, it is found that this approach has a small negative impact on the magnitude of forecast errors relative to MOS with a decrease in the number of very good forecasts, but a large positive impact on the variance about the forecast errors: forecast busts are reduced. As a further step, a second round of MOS is performed on the new model forecasts in a manner identical to the original MOS approach. This second application of MOS results in a significant reduction in both the forecast errors and the variance about those errors relative to the first application of MOS, and includes an increase in very good forecasts.

KEYWORDS: Data assimilation Forecasting Model error

1. INTRODUCTION

The utility of numerical weather prediction (NWP) forecasts is limited by errors in initial conditions, boundary conditions and deficiencies in NWP models. While reducing initial-condition error is important, errors in boundary conditions and model specification mean that accurate forecasts need not be generated by accurate initial conditions; the best 24-hour forecast will be produced by an initial condition that differs from the true system state. This work presents a method for reducing forecast errors by accounting for model error in the forecast stage rather than directly reducing initial-condition error.

The exponential impact of initial-condition error is well known. A huge amount of effort has been expended attempting to reduce this type of error through data assimilation (see, for example, Lorenc (1986) and Talagrand (1997) and the references therein). It is only recently that operational centres have attempted to account for model error in the forecast stage explicitly (Houtekamer *et al.* 1996; Buizza *et al.* 1999). Equations that describe the behaviour of global weather systems are well known, but their intrinsic limitations (such as grid-box averaging) combined with their discretization for use in NWP models introduces errors. Further, processes that evolve on spatial scales smaller than the grid spacing of the numerical model must be parametrized. The combination of these effects results in state-dependent, systematic forecast errors.

The Statistical Modeling Branch of the Meteorological Development Laboratory produces guidance for forecasters through Model Output Statistics (MOS). The MOS technique (Glahn and Lowry 1972) statistically post-processes NWP forecasts to, amongst other things, correct partially for systematic, state dependent model bias. The technique involves determining a linear relationship between predictors (NWP forecasts, previous observations, and climatological information) and predictands (observed weather elements such as min/max temperature, probability of precipitation,

* Corresponding author: Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA 02139, USA. e-mail: jhansen@mit.edu

etc.). Different relationships are derived for different weather elements and different seasons. MOS forecasts are produced for a number of different NWP models, including ensemble forecast products. The MOS produced by the Statistical Modeling Branch has a significant down-scaling component. All MOS down-scales from a grid-point value to a station value and, at the extreme, some synoptically-conditioned subgrid-scale events (e.g. sea breezes) are recovered by the procedure. To date, MOS forecasts are used only by forecasting groups; the information contained in the MOS forecasts does not feed back into model or initial-condition improvement.

This work aims to exploit MOS forecasts in order to improve initial conditions. Here 'improve' does not mean 'move closer to truth', but rather 'result in better forecasts'. When model error exists, initial conditions that are close to the true system state need not result in forecasts that are close to the true system state. Whereas MOS picks up and moves model forecasts closer to the true system state, the procedure described below aims to use MOS states to identify trajectories on the model attractor that are close to the true trajectory on the system attractor. Because the model attractor differs from the system attractor, one may have to move model initial conditions far from the true initial conditions in order to get closer to future forecast states. The procedure used to exploit MOS in this way is described in the following paragraph.

Consider an imperfect-model scenario. An estimate of the true system state, an analysis, is produced through the application of four-dimensional variational data assimilation (4D-Var) (Talagrand 1997). Despite the known model error, 4D-Var is applied as though the model were perfect. The model used here is simple enough to facilitate a proper treatment of model error in 4D-Var, but we wish to mimic the conditions of operational NWP 4D-Var implementations. Forecasts resulting from the sub-optimal 4D-Var initial conditions are processed using the MOS procedure to obtain improved estimates of the future states. A new method for producing analyses is proposed that treats the MOS forecasts as system observations and utilizes 4D-Var to find the initial conditions that bring model forecast states closest to MOS forecast states. Forecast 4D-Var has little impact on root-mean-square (RMS) forecast error relative to the MOS error statistics, but the variance about the RMS forecast error is significantly reduced. A new, independent application of MOS on the forecast 4D-Var states reduces *both* the RMS forecast errors and the variance about those errors relative to the original application of MOS.

The system of equations used in this work is introduced in section 2. The traditional approach to 4D-Var and the associated forecast-error statistics are given in section 3, while MOS and the improvements provided by MOS are given in section 4. Section 5 demonstrates the ability of forecast 4D-Var to reduce forecast-error variance when MOS forecasts are utilized as future 'observations' of the system. The impact of a second round of MOS is shown in section 6. A direct comparison of the results is carried out in section 7, the sensitivity of these results is discussed in section 8, and finally a statement of conclusions is given in section 9.

2. The model

Utilizing MOS forecasts as pseudo-observations of a system's future state is explored in the context of a simple system. In the present study, MOS is used as a forecast post-processing step intended to account for model error; the down-scaling aspect of the operational approach is ignored. In an effort to include 'realistic' model error, sets of equations introduced by Lorenz (1995) are utilized:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F - \frac{h_x c}{b}\sum_{j=1}^{J} y_{j,i}, \quad i = 1, m$$
$$\frac{\mathrm{d}y_{j,i}}{\mathrm{d}t} = -cby_{j+1,i}y_{j+2,i} + cby_{j-1,i}y_{j+1,i} - cy_{j,i} + \frac{h_y c}{b}x_i. \tag{1}$$

Developed as a toy model of an atmospheric quantity distributed zonally about the globe, the equations contain the basic dynamic properties of advection, damping, and external forcing. The equations are cyclic in both x ($x_{m+1} = x_1$) and y ($y_{J+1,m+1} = y_{1,1}$). In this work the external forcing F is set to $F = 8$, the coupling coefficients are set to $h_x = h_y = 1$, and the scaling parameters are given values of $c = 10$ and $b = 10$. There are $m = 18$ equations in x and $J = 5$ equations in y for each x, resulting in a 108-dimensional system. The x-components operate on a spatial scale 10 times larger and a time-scale 10 times slower than the y-components. The J small-scale components coupled to each large-scale component can be thought of as deterministic, perhaps convective, forcing on the large-scale that are in turn forced by the large-scale components. The behaviour of this system is analysed in detail by Hansen (1998).

The large-scale components of Eq. (1) are modelled using

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + \widetilde{F}_i, \quad i = 1, m, \tag{2}$$

where $m = 18$. The form of Eq. (2) is identical to the large-scale components of Eq. (1), but the small-scale components are missing. It is necessary to parametrize the impact of the $F - (h_x c)/b \sum_{j=1}^{J} y_{j,i}$ terms in Eq. (1) in the $\widetilde{F}_i$ term of Eq. (2). A zeroth-order parametrization is performed by specifying all $\widetilde{F}_i$ to take on the constant value of the mean of $F - (h_x c)/b \sum_{j=1}^{J} y_{j,i}$ as determined from a long, independent integration of the system equations. A discussion of Eq. (2) can be found by Lorenz and Emanuel (1998).

The experiments performed in this work specify Eq. (1), hereafter denoted the system, to be 'truth'. The system's large-scale components, the x_i, are imperfectly observed and assimilated into Eq. (2), hereafter denoted the 'model'. Model forecasts are launched from the resulting state estimate, and model-predicted states are compared with system states. Both the system and model are integrated using a fourth-order Runge–Kutta scheme. The system uses a step size of 0.005 while the model is integrated with a step size of 0.05 (0.05 model seconds correspond to approximately 6 hours based on an atmospheric-error doubling time of around 2.5 days). Note that the different integration step sizes, chosen for computational convenience, will introduce an additional source of model error.

3. Four-dimensional variational assimilation

State estimation, or data assimilation, is a crucial component in any forecasting system; without accurate estimates of the true system state it is difficult to produce accurate forecasts of the state's future evolution. Typical approaches to data assimilation involve the combination of a first guess at the system state (typically in the form of a short-term forecast) and system observations weighted by their associated uncertainty. The geophysical literature is rife with data-assimilation studies, and we do not intend to

TABLE 1. DEFINITIONS OF VARIABLES USED IN EQS. (3), (5), (6) AND (8)

Variable	Definition
$\mathbf{x}_{4d1}$	States produced by the first, traditional, application of 4D-Var. These are both past states (over the assimilation window) and forecast states
$\mathbf{x}_{MOS1}$	Forecast states produced by applying MOS to the forecast $\mathbf{x}_{4d1}$ states
$\mathbf{x}_{4d2}$	Forecast 4D-Var states produced by using $\mathbf{x}_{MOS1}$ states as future 'observations' All $\mathbf{x}_{4d2}$ are forecast states; some are over the forecast 4D-Var assimilation window, and others are free forecasts
$\mathbf{x}_{MOS2}$	Forecast states produced by applying a second, independent round of MOS to the forecast 4D-Var states, $\mathbf{x}_{4d2}$
$\mathbf{B}_1$	The background-error covariance used in the traditional application of 4D-Var to produce $\mathbf{x}_{4d1}$ states
$\mathbf{B}_2$	The background-error covariance used in forecast 4D-Var to produce $\mathbf{x}_{4d2}$ states
$\mathbf{R}_i$	The observational-error covariance used in the traditional application of 4D-Var to produce $\mathbf{x}_{4d1}$ states
$\mathbf{R}_{MOS,i}$	The observational-error covariance used in forecast 4D-Var to produce $\mathbf{x}_{4d2}$ states

Notation follows Ide *et al.* (1997) when possible, but the non-traditional use of 4D-Var requires several modifications.

produce new insight into the 4D-Var approach; we will simply utilize 4D-Var as a tool to produce improved estimates of the system state.

The aim of 4D-Var is to generate a model state at the beginning of a trajectory of past observations that, when propagated forward, comes as close as possible to each of the observations in the assimilation 'window'. This initial state is determined through the minimization of a cost function. The cost function is based on the misfit between the model trajectory and associated observations. The misfit between the first guess at the initial state and estimates of the initial state is also included. The cost function has the form

$$\mathcal{J}\{\mathbf{x}_{4d1}(t_{-n})\} = \frac{1}{2}\{\mathbf{x}_{4d1}(t_{-n}) - \mathbf{x}^{b}(t_{-n})\}^{T}\mathbf{B}_1^{-1}\{\mathbf{x}_{4d1}(t_{-n}) - \mathbf{x}^{b}(t_{-n})\} + \frac{1}{2}\sum_{i=-n}^{0}[\mathbf{H}_i\{\mathbf{x}_{4d1}(t_i)\} - \mathbf{y}_i^{o}]^{T}\mathbf{R}_i^{-1}[\mathbf{H}_i\{\mathbf{x}_{4d1}(t_i)\} - \mathbf{y}_i^{o}], \tag{3}$$

for an assimilation window of length n observation times. The first term in Eq. (3) represents the misfit between the best estimate of the system state at the *beginning* of the n-step assimilation window, $\mathbf{x}_{4d1}(t_{-n})$, and a first guess at the best estimate, $\mathbf{x}^{b}(t_{-n})$, weighted by the inverse of the first-guess error covariance, $\mathbf{B}_1^{-1}$; superscript T is the transpose. To start the data-assimilation system, the original first-guess is selected randomly. Subsequent first guesses are provided by short-term forecasts from the analysed states. The second term takes into account the misfit between the best estimate of the system state at a particular time, $\mathbf{x}_{4d1}(t_i)$ ($\mathbf{H}_i$ is a linear operator that maps from model space to observation space), and the corresponding observation, $\mathbf{y}_i^{o}$. Each distance is weighted by the inverse of the observational-error covariance, $\mathbf{R}_i^{-1}$. See Table 1 for a description of the important variables used in this work.

In the following experiments, the observation operator, $\mathbf{H}_i$, is the identity matrix (all $m = 18$ components are observed) and observations are drawn from the system with a component-wise expected observational error of 0.2 (this is 2.5% of the absolute range of values taken on by each component). For simplicity, both the background-error variance and the observational-error variance are assumed to be the same. The structure of the observational-error covariance is diagonal, while the structure of the background-error covariance is consistent with the expected covariance between model components. This covariance is calculated from a large sample of independent realizations drawn

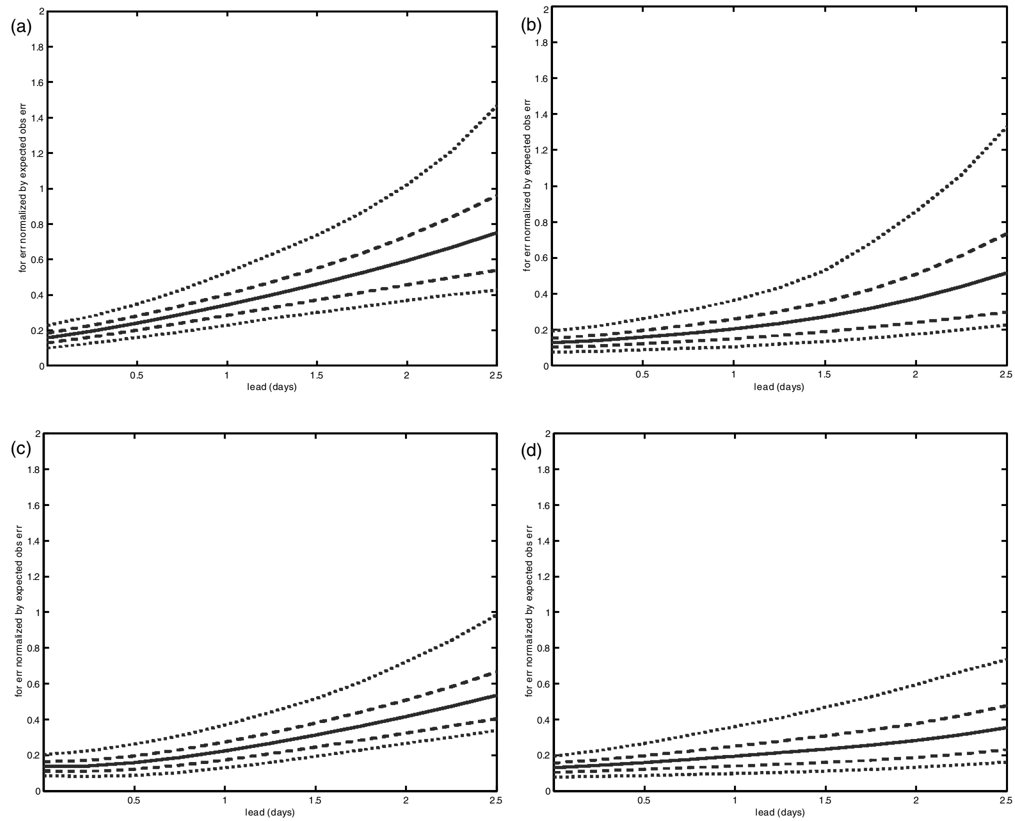

Figure 1. Forecast errors as a function of lead time for (a) raw forecasts, (b) the original application of Model Output Statistics (MOS), (c) forecast 4D-Var, and (d) the second application of MOS. All errors are normalized by the expected observational error. In each panel the full line is the mean, the dashed lines represent one standard deviation about the mean, and the dotted lines are 1st and 99th percentiles. Errors are the 18-dimensional distance from the forecast state to the true state, and statistics were generated using 4096 forecasts.

from the model attractor. Rather than being an estimate of the uncertainty structure associated with the first guess, this covariance is intended to provide information about the way in which the model components are related. This approach is typically taken in geophysical applications of 4D-Var; the aim is to spread the influence of observations in a dynamically consistent way. The assimilation window is set to $n = 4$ (five observation times, or 24 hours). Predictions are run 2.5 days into the future and error statistics are collected for each lead time.

Forecast errors are defined as the Euclidean distance between the forecast state and the true state in the full 18-dimensional space. Statistics are generated over 4096 different forecasts, and are plotted in Fig. 1(a). Forecast error normalized by the expected observational error is plotted as a function of forecast lead time.

4. Model Output Statistics I

In the operational MOS approach, linear relationships are derived that relate verifications at specified lead times to model forecasts. Introduced about 30 years ago (Glahn and Lowry 1972), MOS forecasts continue to provide excellent guidance for operational weather forecasters. The current operational implementation uses NWP model forecasts, available observations, and climatological information as predictors to produce improved values of products such as minimum and maximum temperature, dew-point, wind speed and direction, and precipitation information.

In the present work, the only predictors are model forecast states. Each lead time requires its own linear relationship; the linear relationship between one-step forecasts and verification will be different from the linear relationship between five-step forecasts and verification. The relationships are determined by collecting a large number of forecasts for the same lead time and the corresponding verifications. The mean is removed from each, and linear regression is employed to produce the best (in the least-squares sense) linear relationship between predictors and predictands. For the sake of clarity, each time 'forecast' or 'verification' is mentioned in the discussions of MOS construction, forecast anomaly and verification anomaly are implied.

Let $\mathbf{X}_{4\mathrm{d}1}(t)$ be a matrix whose column vectors are forecasts from a large number of different 4D-Var analyses all valid at a lead time of t. The verifications associated with each of the forecasts in $\mathbf{X}_{4\mathrm{d}1}(t)$ make up the column vectors in $\mathbf{X}_{\mathrm{ver}}(t)$. One wants the linear operator, $\mathbf{A}_{\mathrm{MOS1}}(t)$, such that

$$\mathbf{X}_{\mathrm{ver}}(t) - \mathbf{A}_{\mathrm{MOS1}}(t) f(\mathbf{X}_{4\mathrm{d}1}(t), t), \tag{4}$$

is minimized in a least-squares sense. Here $f(\mathbf{X}_{4\mathrm{d}1}(t), t)$ denotes the predictors to be some function of $\mathbf{X}_{4\mathrm{d}1}(t)$ and lead time. Once $\mathbf{A}_{\mathrm{MOS1}}(t)$ is determined, then the MOS forecast for lead time t, $\mathbf{x}_{\mathrm{MOS1}}(t)$, is obtained by

$$\mathbf{x}_{\mathrm{MOS1}}(t) = \mathbf{A}_{\mathrm{MOS1}}(t) f(\mathbf{x}_{4\mathrm{d}1}(t), t) \tag{5}$$

where $\mathbf{x}_{4\mathrm{d}1}(t)$ is a single forecast state.

All $m = 18$ components of the model forecast were used to predict all $m = 18$ components of the verification. For a lead time of $t = 0$, the predictors were simply the analysis, $f(\mathbf{x}_{4\mathrm{d}1}(t), t) = \mathbf{x}_{4\mathrm{d}1}(t)$. For a lead time of $t = 1$, both the forecast state at $t = 1$ and the analysis were used to predict the verification at $t = 1$, $f(\mathbf{x}_{4\mathrm{d}1}(t), t) = \mathbf{x}_{4\mathrm{d}1}(0) : \mathbf{x}_{4\mathrm{d}1}(1)$. For all verification times greater than $t = 1$ the forecasts valid at t, $t - 1$ and $t - 2$ are used as predictors, $f(\mathbf{x}_{4\mathrm{d}1}(t), t) = \mathbf{x}_{4\mathrm{d}1}(t-2) : \mathbf{x}_{4\mathrm{d}1}(t-1) : \mathbf{x}_{4\mathrm{d}1}(t)$. Predictors were chosen on the basis of convenience and empirical results. The aim was to produce good MOS forecasts rather than attempt to mimic the operational MOS procedure. There were 4096 realizations used in the construction of the MOS operators.

The verification data used in the construction of MOS operators come from 4D-Var analyses drawn from the middle of the assimilation window; a 're-analysis'-like product. The 4D-Var approach exploits model dynamics in such a way that the model trajectory becomes closest to truth in the middle of the assimilation window (Farmer and Sidorowich 1991; Pires *et al.* 1996) making it the best estimate of the full system state available. This is a state estimate that is available to operational NWP centres.

Out-of-sample MOS forecast-error information is shown in Fig. 1(b). As before, forecast error normalized by the expected observational error is plotted as a function of forecast lead time. Notice the significant improvement over the unprocessed forecasts in Fig. 1(a).

For this combination of model and system, the improvement in MOS forecasts comes from MOS's ability to correct the model's tendency to overestimate system values. The parametrization of the impact of the system's small-scale components results in model variability that is larger than system variability. MOS proves adept at accounting for this type of error. A discussion of the more geophysical case where system variability is greater than model variability, and of MOS's natural tendency to reduce forecast variance, is presented in section 8; it will be shown that the results presented in this and the following two sections are robust.

5. Forecast 4D-Var

The fact that the MOS forecast errors and error statistics differ from those of the raw model suggests that MOS forecasts may contain exploitable information. This idea is put to the test by treating MOS forecasts as 'observations' of the future state of the system in a 4D-Var framework.

Model states at $t = 0$, the analysis time, are adjusted to minimize the cost function:

$$\mathcal{J}\{\mathbf{x}_{4\text{d}2}(t_0)\} = \frac{1}{2}\{\mathbf{x}_{4\text{d}2}(t_0) - \mathbf{x}_{4\text{d}1}(t_0)\}^{\text{T}}\mathbf{B}_2^{-1}\{\mathbf{x}_{4\text{d}2}(t_0) - \mathbf{x}_{4\text{d}1}(t_0)\} + \frac{1}{2}\sum_{i=0}^{n}[\mathbf{H}_i\{\mathbf{x}_{4\text{d}2}(t_i)\} - \mathbf{x}_{\text{MOS}1}(t_i)]^{\text{T}}\mathbf{R}_{\text{MOS},i}^{-1}[\mathbf{H}_i\{\mathbf{x}_{4\text{d}2}(t_i)\} - \mathbf{x}_{\text{MOS}1}(t_i)]. \quad (6)$$

The first guess is the original 4D-Var analysis, $\mathbf{x}_{4\text{d}1}(t_0)$, and the first-guess weighting uses the same covariance structure and magnitude as the background-error covariance in Eq. (3). Large amounts of effort are expended on constructing first-guess weightings at operational NWP centres, and rather than asking operational NWP centres to build a new first-guess weighting, we explore the possibility of reusing the existing weighting. However, one can imagine constructing a first-guess weighting based on the difference between states at the end of the assimilation window, $\mathbf{x}_{4\text{d}1}(t_0)$, and the states in the middle of the assimilation window valid at the same time. Such a weighting function was found to have a structure similar to that of $\mathbf{B}_1$, but with lower levels of uncertainty. The impact of this background covariance was a slight improvement over the forecast 4D-Var results shown below, but did not qualitatively change the overall results. The second term in Eq. (6) calculates the misfit between model-produced forecasts and MOS forecasts. The weighting term, $\mathbf{R}_{\text{MOS},i}^{-1}$, is the inverse of the MOS error covariance for each lead time determined by the differences between MOS forecasts and the associated mid-assimilation window analyses. The formulation in Eq. (6) assumes that the MOS errors are uncorrelated in time. This is unlikely to be true. Attempting to account for the temporal correlation in the cost function would make the subsequent minimization operationally impossible given the current situation in NWP centres. We choose to take the pragmatic approach of assuming the MOS errors are temporally uncorrelated, allowing the MOS observational covariances to be applied in a manner similar to the background-error covariance, $\mathbf{B}_2$. The resulting forecast 4D-Var predictions will provide an upper bound on forecast errors; proper treatment of the temporal correlation between MOS forecasts will only improve the forecast 4D-Var states. As in section 3, the assimilation window length is one model day, $n = 4$. It is important to reiterate that this is *forecast* 4D-Var; there are no observations available over the assimilation window because the assimilation window extends into the future. MOS forecasts are being treated as pseudo-observations.

Forecast 4D-Var results are presented in Fig. 1(c). The mean forecast errors are slightly larger than the MOS forecast errors shown in Fig. 1(b), with a noticeable increase in the analysis error. This is consistent with the idea that improved forecasts need not be associated with initial conditions that are closer to truth. What is striking about the forecast 4D-Var results is the reduction in the variance about the mean errors. The extent of the standard deviation bound and the ninety-ninth percentile have both been reduced significantly relative to the MOS results. It is true that this means that the number of very good forecast results has diminished, but with the desirable trade-off that the magnitude of forecast busts has been reduced.

The improvement comes in two forms: (1) from finding initial conditions that produce the lower variance states demanded by the MOS forecasts, and (2) forecast 4D-Var behaves like a dynamical filter to correct MOS forecasts that are physically unrealizable. Forecast 4D-Var predictions are consistent with model dynamics, and as such a re-introduction of model error will have taken place. Further improvements can be expected by a second application of post-processing to remove systematic model error once again.

6. Model Output Statistics II

The dynamic filtering applied to MOS forecasts by forecast 4D-Var serves to reduce the forecast-error variance, but at the expense of a slight increase in the magnitude of forecast errors. The application of forecast 4D-Var will re-introduce systematic model error that was partially accounted for by the MOS procedure. This model error can be reduced by a second, independent application of the MOS framework.

The application is identical to that described in section 4 except that forecast 4D-Var states, $\mathbf{X}_{4d2}(t)$, replace the original 4D-Var forecasts, $\mathbf{X}_{4d1}(t)$. One wants the linear operator, $\mathbf{A}_{MOS2}(t)$, such that

$$\mathbf{X}_{ver}(t) - \mathbf{A}_{MOS2}(t) f(\mathbf{X}_{4d2}(t), t), \tag{7}$$

is minimized in a least-squares sense. Again $f(\mathbf{X}_{4d2}(t), t)$ denotes the predictors to be some function of $\mathbf{X}_{4d2}(t)$ and lead time. Once $\mathbf{A}_{MOS2}(t)$ is determined, then the new MOS forecast for lead time t, $\mathbf{x}_{MOS2}(t)$, is obtained by

$$\mathbf{x}_{MOS2}(t) = \mathbf{A}_{MOS2}(t) f(\mathbf{x}_{4d2}(t), t) \tag{8}$$

where $\mathbf{x}_{4d2}(t)$ is a single forecast state.

Results for the second application of MOS (MOS2) are presented in Fig. 1(d). Improvements over the forecast 4D-Var results (Fig. 1(c)) and the first application of MOS (Fig. 1(b)) are striking. Mean MOS2 forecast errors are significantly reduced relative to the original MOS forecast errors, and the variance about those errors is similar to the forecast 4D-Var variances; the second application of MOS reduces both the expected magnitude of the forecast errors and the size of forecast busts. Unlike forecast 4D-Var, the MOS2 does not suffer from a reduction of very good forecasts: MOS2 results in an increase in very good forecasts. It is natural to wonder whether this improvement can continue with repeated applications of forecast 4D-Var and MOS. For the system investigated, further applications do not produce improved forecast results. This suggests that the first application has accounted for a maximal amount of linear model error.

7. Comparison of results

The improvements realized by MOS2 forecasts are shown in Fig. 2. Plotted in each panel is the factor of improvement realized by MOS2 relative to raw forecasts, the first application of MOS and to forecast 4D-Var. The improvement factor is defined as the ratio of raw, MOS or forecast 4D-Var forecast statistic to the associated MOS2 forecast statistic. A value less than one means MOS2 degrades the forecast statistic while a value greater than one means MOS2 improves the forecast statistic (e.g. an improvement

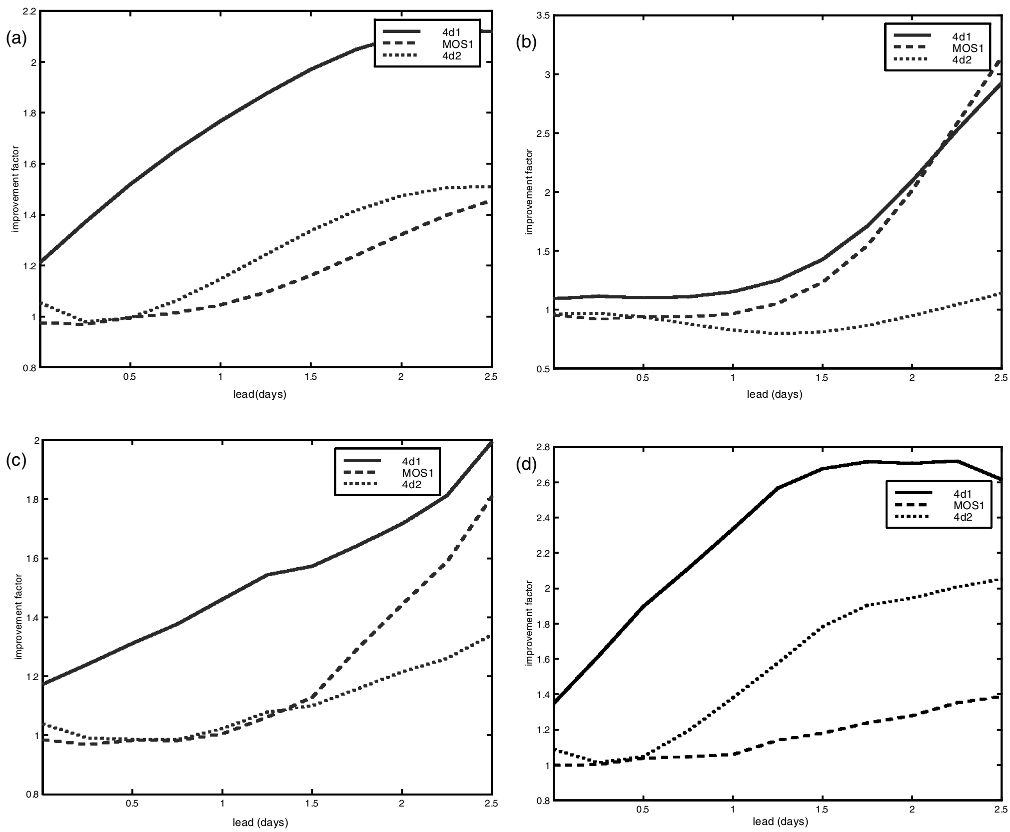

Figure 2. Improvement factors relative to MOS2 forecast statistics (see text): (a) the root-mean-square improvement in the mean forecast errors, (b) the improvement in the variance about those errors, (c) the improvement in forecast busts, and (d) the improvement in the best forecasts (the 1st percentile of the error distribution). In each panel the full line represents improvements relative to the raw forecasts, the dashed line is improvements relative to the first application of MOS, and the dotted line is improvements relative to forecast 4D-Var.

factor of two implies MOS2 forecasts are twice as good as the forecast with which it is being compared). The different panels represent different statistics.

Figure 2(a) demonstrates that the MOS2 mean forecast errors are always superior to the raw forecast and to the forecast 4D-Var results, but the original MOS analysis and six-hour forecast is slightly better than the MOS2 results. This plot also shows that the forecast 4D-Var mean analysis error is worse than the original MOS analysis error. To get close to MOS forecasts, model error required the analysis to move away from the true initial condition.

The MOS2 reduction in variance about the mean forecast error shown in Fig. 2(b) is striking when compared with the raw forecasts and the original MOS forecasts. The MOS2 variance results are nearly identical to the forecast 4D-Var variance results, suggesting that the improvement in variance comes from the dynamic filtering aspect of forecast 4D-Var.

The forecast bust improvement due to MOS2 is shown in Fig. 2(c). Forecast busts are defined here by the ninety-ninth percentile of the forecast-error distribution. MOS2 reduces forecast busts relative to raw forecasts, the first application of MOS and forecast 4D-Var especially at the longer leads. It is interesting to note the improvement in the extremes of the error distribution of MOS2 relative to forecast 4D-Var despite the equivalence of their variance statistics (see Fig. 2(b)). The MOS2 processing is able to account for the systematic overestimation of the system's extreme events. The relatively small number of extreme events means that this improvement is not observed in the variance comparison.

The MOS2 improvement in the level of very good forecasts is shown in Fig. 2(d). Very good forecasts are defined here by the first percentile of the forecast-error distribution. MOS2 improves the level of very good forecasts relative to raw forecasts at all leads. There is relatively little improvement in very good forecasts relative to the first application of MOS and forecast 4D-Var over the first 6 hours. The improvement in the very good forecasts produced by MOS2 over those produced by the first application of MOS is relatively small; only 20% at a lead of 2.5 days. The second application of MOS proves capable of correcting the degradation in very good forecasts produced by forecast 4D-Var, suggesting the degradation found in forecast 4D-Var is the result of model inadequacies.

8. Sensitivity of results

The results presented in this work all benefit from complete observations. At every observation time all $m = 18$ components were observed. Operational data-assimilation schemes do not have this luxury, so experiments to assess the sensitivity of results to the number of observed components were carried out. In general, reducing the number of observed components increases the analysis error. So long as the combination of the number of observed components and the structure of the observational network (the pattern of components observed) results in analysis errors that are small enough to allow the model error signal to be observed, the forecast 4D-Var results are robust. As the analysis error increases, little to no improvement is observed in the short lead-time forecasts, but improvements are still made in the long-lead forecasts. Again, this is because at longer leads the model error signal is discernible from error due to initial condition mis-specification.

Returning to the case where all components are observed, one can compare the MOS relationships between the first application of MOS and the second application of MOS. The linear relationships determined for the first application of MOS are found to be different from those found for MOS2. Attempting to correct forecast 4D-Var states using the original MOS linear relationships degrades forecasts. The degradation is severe at short lead-times, but forecast states at longer leads are only slightly worse than MOS2 forecasts. The reduction of the model error signal in forecast 4D-Var states, especially over the length of the assimilation window, renders the original MOS relationships inappropriate. The model error signal in forecast 4D-Var has the opportunity to re-emerge at longer leads, so the application of the original MOS relationships is justifiable. This result suggests there may be scope for testing whether the second application of MOS is likely to be successful without having to collect a large database of forecast 4D-Var states and verifications.

If the second application of MOS did not improve the very good forecasts relative to forecast 4D-Var, one could conclude that the improvement in the mean RMS error, the variance in RMS error, and the forecast busts were simply due to MOS and forecast 4D-Var 'regressing to the mean'. This is not the case here. One can examine the ratio of the variance* of raw, MOS, forecast 4D-Var, and MOS2 forecast fields to the variance

* The variance in the 18-dimensional space.

of the true field. The variance in the fields from the first application of MOS stay within 3% of the variance of the true field (on average) at all leads out to 2.5 days (figure not shown). The variance in the fields from the second application of MOS stay within 1% of the variance of the true field at all leads out to 2.5 days. The variance in the fields from the raw forecasts and from forecast 4D-Var both grow relative to the variance in the true field. At analysis time both are within 4% of the true variance, but by a lead of 2.5 days they have grown to a variance about 11% *larger* than the true variance. It is clear that the improvement in error statistics is not coming from a reduction in forecast-field variance. Two further tests were performed to ensure that a relaxation to climatology was not responsible for the results presented in this work. The first was simply to calculate forecast-error statistics when the forecast is constrained to be climatology. The mean normalized error in this experiment was 4.8, which should be compared with the normalized forecast-error statistics presented in Figs. 1(a)–(d) where the largest value seen was less than 1.5. The second experiment was to construct an MOS relationship with only two predictors: the model forecast field, and the climatological field. Even at leads of 2.5 days the model forecast field received a weighting of over 0.95 compared with a weighting of less than 0.05 for climatology. This simple application of MOS was not able to match the performance of the more complex MOS strategies presented earlier.

The structure of Eq. (1) results in the variance of the large-scale components of the system being smaller than the variance of the fields produced by the model, Eq. (2). This is not what is typically found in NWP, where the model variance is smaller than the observed variance. To test whether the results presented in this work depend on the model variance being larger than the system variance, Eq. (1) was altered by replacing the minus sign before the $h_x c/b$ term with a plus sign. The result of this change is to increase the variance of the system's large-scale components to a level above what is produced by the model, Eq. (2). While still a nonlinear dynamical system, the modified system equations no longer conserve energy (but they do not blow up) . Repeating the experiments of the previous sections using the modified system did not qualitatively change the results reported above, and even in the case where the model variance is deficient, the improvement given by forecast 4D-Var and the second application of MOS is not a result of a relaxation to climatology. Of course, if one considered forecast leads many error doubling times into the future, it is not expected that MOS would provide information beyond what is available in climatology.

9. Conclusions

Post-processing of model forecasts using the MOS procedure results in future state estimates that are superior to raw model forecasts. Treating the MOS forecasts as pseudo-observations of the future state of the system and applying a forecast 4D-Var technique (an assimilation window that extends into the future) results in mean forecast errors similar to the MOS errors, but a reduction in the variance about these errors. The use of the imperfect model in forecast 4D-Var re-introduces systematic model error that can be accounted for by a second, independent application of the MOS procedure. In the current work, the forecast 4D-Var cost function assumes that the MOS errors are uncorrelated in time. This was done to mimic the operational NWP setting where inclusion of the temporal correlation would cause the problem to become computationally not feasible. Inclusion of the temporal correlation would likely improve the forecast 4D-Var results and reduce the high level of improvement seen in the second application of MOS.

The MOS approach is able to account for systematic biases in component variance. If the variance of model components is systematically too large or too small then MOS should be able to produce improved estimates of the system state. Forecast 4D-Var then finds initial conditions that force the model to be as close as possible to these improved estimates while filtering any unrealizable MOS states resulting from the statistical correction process.

One would not expect forecast 4D-Var to be effective in situations where MOS adds little value and/or where the model is extremely deficient. Linear MOS would add little value if the structure of the model error were highly nonlinear or if the analysis errors are so large that systematic model errors are obscured. Even if the MOS approach produces improved forecasts, if the model is so poor that it is unable to shadow the future evolution of the system, forecast 4D-Var will have difficulty producing initial conditions that improve forecasts.

Pains were taken to ensure that quantities used in the 4D-Var cost function and in the MOS formulations were operationally obtainable. Covariances and MOS relationships were all static, and all constructed relative to analysed, not true system states. Insofar as 4D-Var is an operationally realizable state estimation technique and MOS is capable of improving model forecasts, it is expected that forecast 4D-Var would be beneficial in operational NWP applications. There would be issues associated with the down-scaling aspect of NWP MOS. Rather than perform down-scaling, MOS valid only at model grid-points would need to be implemented. The utility of this grid-point MOS will be completely dependent upon the quality of NWP analyses; if the forecast error due to a poor analysis is large relative to systematic model error, MOS will not improve the raw forecasts, and forecast 4D-Var is unlikely to provide additional improvement. The accuracy of MOS can also be compromised by changes being made to the model without associated changes to the MOS relationships. To be effective, operational centres would need to keep an archive of observations that would be used to train new MOS relationships each time the model is improved.

The grid-point approach to MOS would only be necessary for the first application of MOS; MOS2 could make full use of down-scaling. Operational MOS predictands in NWP are local in space. While the predictors may have a spatial distribution, the predictand is a single station of interest. A further benefit of forecast 4D-Var would be its ability to spread the influence of MOS forecasts just as standard 4D-Var spreads the influence of isolated observations.

Acknowledgements

The authors gratefully acknowledge discussions with Zoltan Toth and Mark Beuhner regarding the operational applicability of the techniques discussed in this paper. Conversations with Mark Hamill, Ed Lorenz and Chris Snyder contributed greatly to the quality of the work. The authors are also thankful for the comments of three anonymous reviewers. This research was supported under National Science Foundation (NSF) Contract ATM-9815114, NSF Contract ACI-0121182, Cal Tech Jet Propulsion Laboratory (USA) Contract 1205624, and National Aeronautics and Space Administration Contract NAG5-7857.

References

Buizza, R., Miller, M. and Palmer, T. N. 1999 Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125,** 2887–2908

Farmer, J. D. and Sidorowich, J. J. 1991 Optimal shadowing and noise reduction. *Physica D*, **47,** 373–392

Glahn, H. R. and Lowry, D. A. 1972 The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteorol.*, **11,** 1203–1211

Hansen, J. A. 1998 'Adaptive observations in spatially-extended, nonlinear dynamical systems'. Ph.D. Thesis, Oxford University

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996 A system simulation approach to ensemble prediction. *Mon. Weather Rev.*, **124,** 1225–1242

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. 1997 Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Japan*, **75**(1B), 181–189

Lorenc, A. C. 1986 Analysis method for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112,** 1177–1194

Lorenz, E. N. 1995 'Predictability—A problem partly solved'. In: Predictability. ECMWF, Seminar Proceedings, Shinfield Park, Reading, RG2 9AX, UK

Lorenz, E. N. and Emanuel, K. A. 1998 Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55,** 399–414

Pires, C., Vautard, R. and Talagrand, O 1996 On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, **48A,** 96–121

Talagrand, O. 1997 Assimilation of observations, an introduction. *J. Meteorol. Soc. Japan*, **75**(1B), 191–209