Model error estimation for the CPTEC Eta model*

Michael K. Tippett† Arlindo da Silva‡

June 2, 1999

Abstract

Statistical data assimilation systems require the specification of forecast and observation error statistics. Forecast error is due to model imperfections and differences between the initial condition and the actual state of the atmosphere. Practical four-dimensional variational (4D-Var) methods try to fit the forecast state to the observations assuming that model error is negligible. Here with a number of simplifying assumption, a framework is developed for isolating the model error given the forecast error at two lead-times. Two definitions are proposed for the Talagrand ratio τ, the fraction of the forecast error due to model error rather than initial condition error. Data from the CPTEC Eta Model running operationally over South America are used to calculate forecast error statistics and lower bounds for τ.

1 Introduction

Data assimilation systems combine satellite data and other measurements with a first guess coming from a predictive model to produce an analysis or estimate of the state of the atmosphere. This estimate can be used as an initial condition for numerical weather prediction, or a sequence of estimates can be used to study Earth Science phenomena. In statistical data assimilation methods the analysis is a weighted average of the model forecast and current observations. The weighting is determined by the specification of forecast and observation error statistics. For instance, where forecast error is large, more weight is given to observations. Forecast error statistics also determine how observations correct forecast errors in a neighborhood of the observation.

*This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grant 381737/97-7, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Grant 1998/11952-0 and the NASA EOS Interdisciplinary Project on Data Assimilation.

†International Research Institute for climate prediction, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964-8000, USA (tippett@iri.ldgo.columbia.edu). This work was done while the author was with the Centro de Previsão de Tempo e Estudos Climáticos, Cachoeira Paulista, SP, Brazil.

‡Data Assimilation Office, Code 910.3, NASA/GSFC, Greenbelt, MD 20771, USA (dasilva@dao.gsfc.nasa.gov).

Forecast error has two distinct sources: (i) model deficiencies and (ii) inaccurate initial conditions. In principle, the forecast error is found by solving a Fokker-Planck equation (Epstein 1969) or in the absence of model error a Liouville equation (Ehrendorfer 1994). However, such approaches are infeasible for realistic weather prediction models with a large number of degrees of freedom, and approximations are necessary. In the linear-tangent framework, forecast error growth due to initial condition error is characterized by the singular vectors of the dynamics (Farrell 1990, Ehrendorfer & Tribbia 1997). Monte Carlo or ensemble methods provide a means of calculating the nonlinear growth of forecast error due to differences between the model initial condition and the actual state of the atmosphere (Leith 1974, Houtekamer & Mitchell 1998).

Determination of the forecast error due to model imperfections is conceptually considerably more difficult, since the model error is unknown and potentially unknowable (Cohn 1997). Advanced sequential data assimilation methods based on the Kalman filter and its generalizations require that the model error be specified. Misspecification of model error can have a substantial detrimental effect on data assimilation performance (Dee 1995). Variational methods also require the specification of observation and model error statistics (Bennett 1992, Ghil & Malanotte-Rizzoli 1991). Practical adjoint variational methods, like those used operationally, force the model to fit the observations and assume that model error is negligible (**?**). This "perfect model" hypothesis is a weakness of four-dimensional variational (4D-Var) methods and can lead to serious problems (Miller et al. 1994)

In this work we develop a framework for separating the forecast error into components due to error in the initial condition and due to errors in the model. We show that given the forecast error at two lead-times, the model error can be obtained. To arrive at this result requires substantial assumptions and simplifications. First, ensemble averages are replaced by time-averages and the error statistics assumed stationary. In this framework, average properties of the model error can be estimated given bounds on the growth properties of the dynamics. Detailed model error properties can be calculated if the dynamics sufficiently is simple that it can be applied to an error covariance matrix. Here, we use a forecast-anomaly Markov model to approximate the dynamics. Perhaps the most serious assumption is to assume access to the forecast error. Here we approximate forecast error by the difference of forecast and analysis. In the Concluding Remarks we suggest some alternatives.

The method is applied to forecast and analysis data from the CPTEC Eta Model running operationally with a resolution of 40 km over South America (Black 1994, Chou 1996) at 00 UTC and

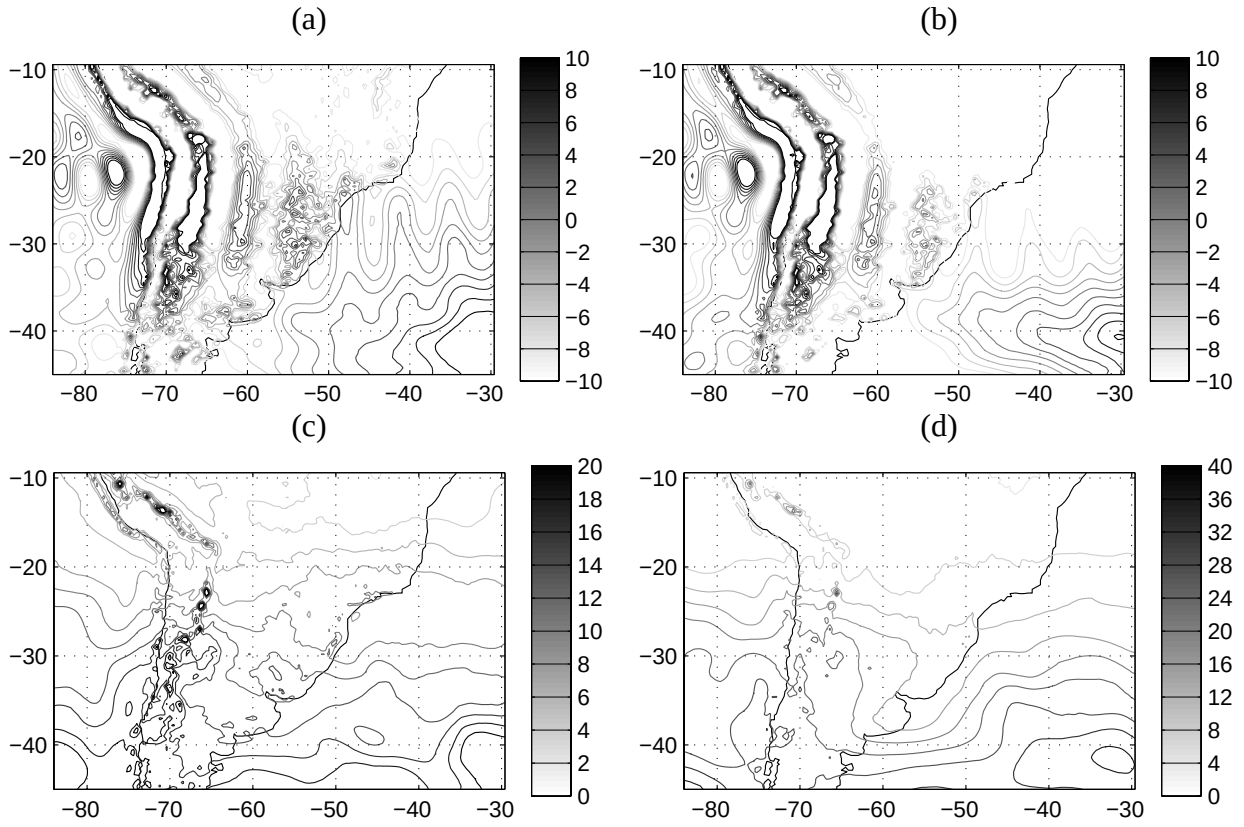

Figure 1: Contour plots at 500 mb of the forecast height error bias at (a) 24 hours (b) 48 hours and of the forecast height error standard deviation at (c) 24 hours and at (d) 48 hours; units are meters.

12 UTC. The CPTEC Eta Model forecast error is approximated by the difference of the forecast and the NCEP operational analysis. We note that quality of the 00 UTC and 12 UTC NCEP analyses is likely different since the number of observations tends to be greater at 12 UTC. The NCEP analysis, interpolated onto the Eta grid, is also the initial condition for the operational CPTEC Eta model. We calculate forecast error statistics for the period during August of 1998 using a subset of the complete set of predicted variables, namely the height fields at three levels 300, 500 and 700 mb on the region 85W to 30W and 45S to 10S interpolated onto a $0.4^\circ \times 0.4^\circ$ grid. This resolution gives a total of $90 \times 137 \times 3$ grid points; all calculations show here are three-dimensional and use this grid. The forecast error mean and standard deviation are shown at the 500 mb level in Fig. 1.

Much of the systematic difference between forecast and analysis is related to the spectral topography representation of the analysis. The spectral topography of the NCEP analysis is very different from the mountain-step coordinate representation of the Eta model near the Andes. As a consequence, the NCEP analysis fields interpolated onto the Eta model grid are incompatable with the Eta model topography forcing. Therefore the Eta model 24-hour and 48-hour forecasts

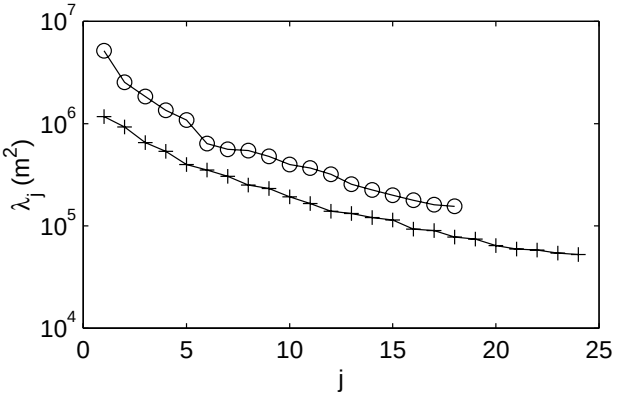

Figure 2: Eigenvalues of the 24-hour (plus signs) and 48-hour (circles) forecast error covariance matrices.

present systematic differences from the NCEP analysis near the Andes. Additionally, the forecast bias contains wave structures. Since this characteristic is less prominent in the 48-hour forecast bias, we hypothesis that it is a residual effect of the incompatability of the initial condition with the Eta topography near the Andes. The bias maxima located in the lower right-hand corner of the domain is likely related to the large natural variability of the height field there. This feature is also present in the forecast error standard deviation. The standard deviation shown has been filtered by performing an eigenvalue decomposition of the forecast error covariance and retaining 90% of the total variance. This means retaining 24 and 18 eigenmodes respectively of the 24-hour and 48-hour forecast error covariances matrices. The eigenvalues of the forecast error covariance matrices at the two lead times are shown in Fig. 2. The 48 hours forecast errors are larger than the 24 hours error. Calculation of the principle angles (see Appendix) between subspaces spanned by the leading eigenvectors of the 24-hour and 48-hour forecast error covariances show that although the first three modes project well onto each other additional modes do not; no simple modal relation is observed between 24-hour and 48-hour forecast errors.

There are two sources of 48-hour forecast errors: (i) error already present in the 24-hour forecast and (ii) deficiencies of the forecast model. Forecast model error includes imperfectly represented physical processes and errors in the boundary conditions. The lateral boundary conditions provided by the CPTEC/COLA GCM are one source of model error (Gustafsson et al. 1998), as are lower boundary conditions (Yang et al. 1994). We would like to know the fractions of the 48-hour forecast error due to propagated 24-hour forecast errors and due to model error. The assumption used in 4D-Var data assimilation methods is that the ratio τ of model error to propagated error is

small. Here we will refer to τ as the Talagrand ratio. In the following section we will suggest some precise definitions for τ.

2 Methodology

In the context of linear dynamics, we can obtain the model error covariance given the forecast error covariances at two lead-times, for instance given the 24-hour and 48-hour forecast error covariances. Define forecast error $\epsilon_{k;k+1}$ to be the error of the forecast starting at time k and valid at time $k+1$. We use time-steps of 24-hours. The 24-hour forecast error $\epsilon_{k;k+1}$ is the difference at time $k+1$ of the 24-hour forecast and the true state w^t_{k+1},

$$\epsilon_{k;k+1} = \mathbf{M}_{k;k+1} w^a_k - w^t_{k+1}\,; \tag{1}$$

w^a_k is the analysis at time k and $\mathbf{M}_{k;k+1}$ is the linear operator that advances the state from time k to time $k+1$. Likewise, define the 48-hour forecast error $\epsilon_{k;k+2}$ to be the difference of the 48-hour forecast and the true state w^t_{k+2} at time $k+2$,

$$\epsilon_{k;k+2} = \mathbf{M}_{k;k+2} w^a_k - w^t_{k+2}\,. \tag{2}$$

The 48-hour forecast error is due to the propagated 24-hour forecast error and the model error as shown by

$$\begin{aligned} \epsilon_{k;k+2} &= \mathbf{M}_{k;k+2} w^a_k - w^t_{k+2} = \mathbf{M}_{k+1;k+2}(\mathbf{M}_{k;k+1} w^a_k - w^t_{k+1}) - \epsilon^t_{k+2} \\ &= \mathbf{M}_{k+1;k+2}\epsilon_{k;k+1} - \epsilon^t_{k+2}\,. \end{aligned} \tag{3}$$

The model error ϵ^t_{k+1} is defined by

$$\epsilon^t_{k+1} = w^t_{k+1} - \mathbf{M}_{k;k+1} w^t_k\,. \tag{4}$$

The 24-hour and 48-hour forecast error covariances $\mathbf{P}_{k;k+1}$ and $\mathbf{P}_{k;k+2}$ satisfy

$$\mathbf{P}_{k;k+2} \equiv \left\langle \epsilon_{k;k+2}\epsilon^T_{k;k+2} \right\rangle = \mathbf{M}_{k+1;k+2}\mathbf{P}_{k;k+1}\mathbf{M}^T_{k+1;k+2} + \mathbf{Q}_{k+1}\,, \tag{5}$$

where we have taken $\left\langle \epsilon^t_k (\epsilon^t_k)^T \right\rangle = \mathbf{Q}_k$; we use the notation $\langle\cdot\rangle$ to denote ensemble average. Therefore, the model error covariance $\mathbf{Q}_{k+1}$ can be obtained given the dynamics $\mathbf{M}_{k+1:k+2}$ and the forecast error covariances $\mathbf{P}_{k;k+1}$ and $\mathbf{P}_{k;k+2}$.

There are difficulties with the method presented above. First, the true state, and hence the forecast error, is unknown. Therefore, we make the approximation of replacing the true state by the verifying analysis, i.e., instead of (1) we take

$$\epsilon_{k;k+1} = \mathbf{M}_{k;k+1} w^a_k - w^a_{k+1} \,. \tag{6}$$

Another difficulty is that even if the true state were known and the forecast error could be calculated, the single realization of the forecast error available at each time-step would not be sufficient to calculate the ensemble averages that define the forecast error covariance matrices. Therefore, we replace ensemble averages by time averages and take the model dynamics and model error covariance to be stationary. In this case, (5) becomes

$$\mathbf{P}_{48} = \mathbf{M}\mathbf{P}_{24}\mathbf{M}^T + \mathbf{Q} \tag{7}$$

where $\mathbf{P}_{24}$ and $\mathbf{P}_{48}$ are respectively the 24-hour and 48-hour forecast error covariances.

The 48-hour forecast error covariance is the sum of the propagated 24-hour forecast error covariance $\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T$ and the model error covariance $\mathbf{Q}$. To calculate these two contributions to the 48 hour forecast error, a linear approximation $\mathbf{M}$ of the dynamics must be applied to the 24 hour forecast error covariance matrix. Such an approximate linear dynamics might come from the tangent linear dynamics linearized about a mean background-state. Here we will use as $\mathbf{M}$ a forecast-anomaly Markov model.

However, before discussing details of this dynamics we suggest some precise definitions for the Talagrand ratio τ and derive some estimates. These definitions and estimates are quite general and depend only on limited information about the forecast error covariance and the linear dynamics. We define the variance Talagrand ratio τ_{var} by

$$\tau_{\text{var}} = \frac{\mathrm{tr}\,\mathbf{Q}}{\mathrm{tr}\,\mathbf{P}_{48}} = 1 - \frac{\mathrm{tr}\,\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T}{\mathrm{tr}\,\mathbf{P}_4 8} \,, \tag{8}$$

i.e., the fraction of the 48 hour forecast error variance due to model error; tr denotes trace. Under the perfect model assumption $\tau_{var} = 0$; on the other hand, when $\mathbf{Q}$ is very large $\tau = 1$. This definition is invariant under orthogonal transformations of the state-space. A lower bound for τ_{var} can be obtained if the total variance of the 24 and 48 hour forecast errors and the maximum possible amplification by the dynamics is known, namely

$$\tau_{\text{var}} \geq 1 - \sigma_1^2(\mathbf{M}) \frac{\mathrm{tr}\,\mathbf{P}_{24}}{\mathrm{tr}\,\mathbf{P}_{48}} \,; \tag{9}$$

$\sigma_1(\mathbf{M})$ is the largest singular value of the 24-hour dynamics $\mathbf{M}$, here the largest factor by which height errors can grow in 24 hours. This lower bound reflects a worst-case situation where all the variance of the 24 hour forecast error projects onto the fastest growing singular vector of the dynamics.

Another lower bound for τ_{var} depends on the variance $\lambda_1(\mathbf{P}_{24})$ associated with the leading eigenvector of the 24 hour forecast error covariance and the quantity $\mathrm{tr}\,\mathbf{M}\mathbf{M}^T$, namely:

$$\tau_{\text{var}} \geq 1 - \frac{\lambda_1(\mathbf{P}_{24})}{\mathrm{tr}\,\mathbf{P}_{48}}\,\mathrm{tr}(\mathbf{M}\mathbf{M}^T)\,. \tag{10}$$

This lower bound reflects a worse-case situation where spectrum of the 24 hour forecast error covariance is flat. Recall that

$$\mathrm{tr}\,\mathbf{M}\mathbf{M}^T = \sum_{i=1}^{n} \sigma_i^2(\mathbf{M})\,, \tag{11}$$

and is the expected amplification of the variance of an uncorrelated, homogeneous random initial condition (Tippett 1999). Therefore, given the singular values of $\mathbf{M}$ and the eigenvalues of the 24 and 48 hour forecast error covariance, lower bounds for τ_{var} can be obtained.

Another measure of the model error is the volume Talagrand ratio τ_{vol} defined as (Schneider & Griffies 1999)

$$\tau_{\text{vol}} = \left(\frac{\det \mathbf{Q}}{\det \mathbf{P}_{48}}\right)^{1/n} = \left(\frac{\det\left(\mathbf{P}_{48} - \mathbf{M}\mathbf{P}_{24}\mathbf{M}^T\right)}{\det \mathbf{P}_{48}}\right)^{1/n}\,; \tag{12}$$

n is the dimension of the state-space. This definition has the following geometric interpretation. For a Gaussian random variable with zero mean and covariance $\mathbf{P}$, the ellipsoid $\mathcal{E}_p(\mathbf{P})$ that encloses some fraction $0 < p < 1$ of the cumulative probability distribution has a volume proportional to $(\det \mathbf{P})^{1/2}$. Therefore, the Talagrand ratio τ_{vol} is the square of the geometric mean of the semiaxis lengths of the model error ellipsoid $\mathcal{E}_p(\mathbf{Q})$ over the square of the geometric mean of the semiaxis lengths of the 48-hour forecast error ellipsoid $\mathcal{E}_p(\mathbf{Q})$. We note that this definition is invariant under general nonsingular transformations of the state-space. As a consequence, τ_{vol} does not depend on the choice of inner product. A lower bound for τ_{vol} is

$$\tau_{\text{vol}} \geq 1 - \left((\det \mathbf{M})^2\,\frac{\det \mathbf{P}_{24}}{\det \mathbf{P}_{48}}\right)^{1/n}\,. \tag{13}$$

Special care in calculating τ_{vol} must be taken when the covariance or dynamics matrices are singular. The simplest remedy and the one we use here is to compute the determinants on a reduced spaces where the matrices are nonsingular.

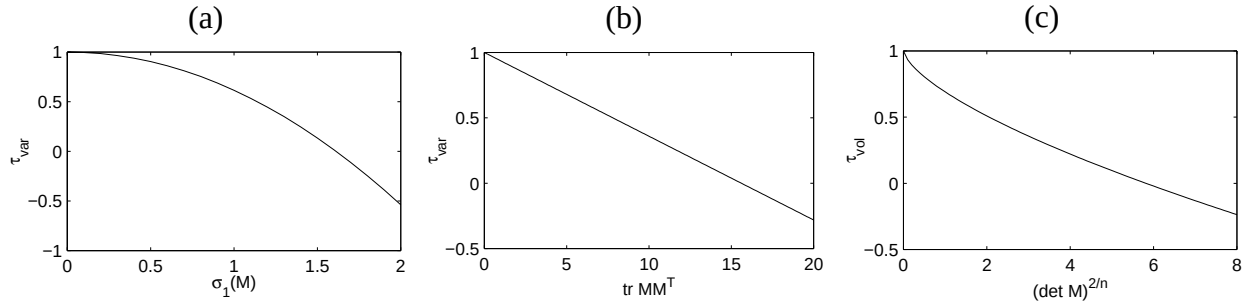

Figure 3: Talagrand ratio τ lower bounds as a function of (a) $\sigma_1(\mathbf{M})$, (b) $\mathrm{tr}\,\mathbf{M}\mathbf{M}^T$ and (c) $(\det \mathbf{M})^{2/n}$.

In Fig. 3 we plot lower bounds for τ_{var} and τ_{vol} as functions of the dynamics for the Eta model forecast error covariance matrices calculated here. If the model dynamics does not amplify the 24-hour forecast error, the 48-hour forecast error is due mainly to the model error $\mathbf{Q}$ and $\tau_{\text{var}} = \tau_{\text{vol}} = 1$. As the amplification ability of the dyanmics is increased, that is as the singular values of the dynamics become larger, the lower bounds for τ_{var} and τ_{vol} decrease; the lower bounds give no information when they are less than zero. We calculate τ_{vol} in a conservative fashion, using only 3 modes of the error covariance matrices; using more modes produces larger lower bounds. Figure 3(a) shows that for modest error growth rates, i.e., ~ 1.5/day the lower bound in (9) gives no information. Figure 3(b) shows that (10) gives a nonzero lower bound for τ if $\mathrm{tr}\,\mathbf{M}\mathbf{M}^T \leq 15$ which is equivalent to random uncorrelated perturbations being amplified by a factor less than 3.87/day. Figure 3(c) shows that (13) gives a nonzero lower bound for τ if the geometric mean of the singular values of $\mathbf{M}$ is less than 2.45/day. These last two results mean that in the absence of relatively strong dynamics to amplify errors, model error is a non-negligible source of forecast error.

3 Results

To investigate the question of dynamical error amplification we calculate a Markov model for the forecast anomaly evolution using data from the same period. The 48-hour forecast anomaly w_{48}^f is assumed to be related to 24-hour forecast anomaly w_{24}^f by

$$w_{48}^f = \mathbf{M} w_{24}^f + b\,, \tag{14}$$

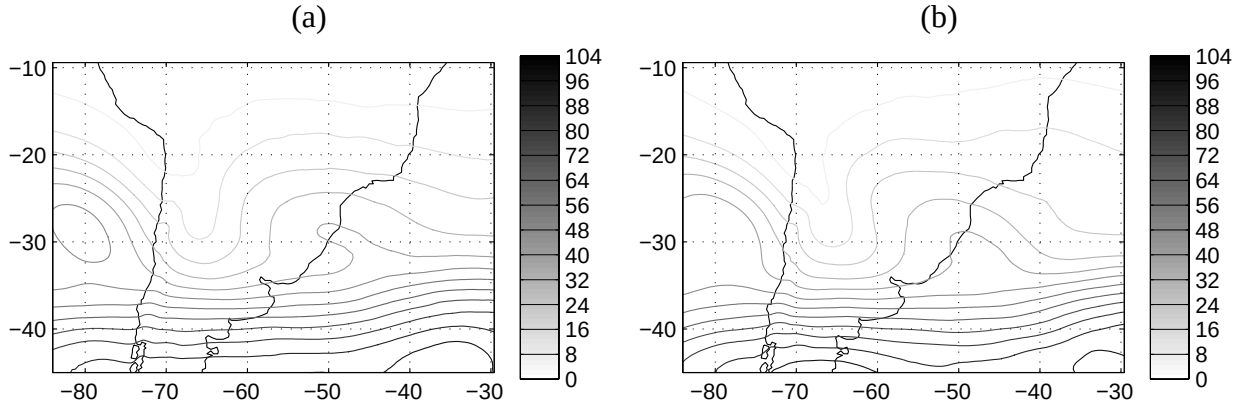

Figure 4: The standard deviations of the (a) 24-hour and (b) 48-hour forecast anomaly height fields plotted at 500 mb. Units are meters

where the residual b is modeled as a random forcing term; the anomolies are defined by subtracting the time-averages. The dynamics $\mathbf{M}$ is found by calculating

$$\mathbf{C} \equiv \left\langle w_{48}^f (w_{24}^f)^T \right\rangle \quad \text{and} \quad \mathbf{P}_{24}^f \equiv \left\langle w_{24}^f (w_{24}^f)^T \right\rangle \tag{15}$$

and solving $\mathbf{C} = \mathbf{M}\mathbf{P}_{24}^f$. The matrix $\mathbf{P}_{24}^f$ is singular since the number of samples is less than the number of degrees of freedom. We deal with this problem by computing the eigenvalue decomposition of $\mathbf{P}_{24}^f$ and setting the trailing eigenvalues to zero. Then we replace $\left(\mathbf{P}_{24}^f\right)^{-1}$ by the Moore-Penrose inverse (Golub & Van Loan 1996). The number of modes to be retained in the calculation is arbitrary. Keeping modes associated with small eigenvalues increases the singular values of $\mathbf{M}$ since retaining more modes makes $\mathbf{P}_{24}^f$ increasingly close to being singular. However, modes associated with small eigenvalues are likely to suffer from sampling error. On the other hand, keeping too few modes produces dynamics that does not produce growth. Here we choose to retain 5 modes of $\mathbf{P}_{24}^f$ explaining 85% of its total variance. Later we show how the lower bounds for τ depend on the number of modes retained.

Figure 4 shows the standard deviation of the 24 and 48 hour forecast height anomaly fields. As expected there are only slight differences between the standard deviation of the 24 and 48 hour height anomaly fields; both should approximate the natural variability of the height field during this period. The variability of the model shown in Fig. 4 can be different from that of the model error shown in Fig. 1. For instance, model error may be a result of the model not presenting the same variability observed in nature. However, it is reasonable to expect model error variablity to have some relation with model error variability and there are some similarities between the two here though the model error has much more small-scale structure. Calculation of the principle

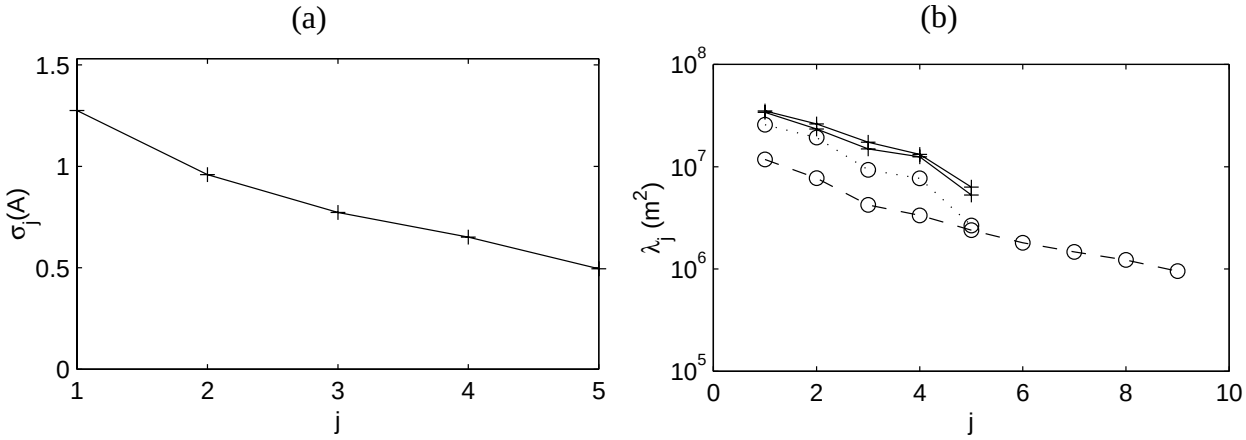

Figure 5: Panel (a) shows the singular values of $\mathbf{M}$. Panel (b) shows the eigenvalues of the forecast anomalies covariances $\mathbf{P}^f_{48}$ and $\mathbf{P}^f_{24}$ (solid lines with plus signs), eigenvalues of $\mathbf{M}\mathbf{P}^f_{24}\mathbf{M}^T$ (dotted line with circles) and eigenvalues of the residual $\left\langle bb^T \right\rangle$ (dashed line with circles).

angles show that the leading eigenmodes of $\mathbf{P}^f_{24}$ and $\mathbf{P}^f_{48}$ span approximately the same subspaces.

Figure 5(a) shows the singular values of $\mathbf{M}$. Although the dynamics is stable by construction with all eigenvalues inside the unit circle, there are singular values greater than one, indicating nonmodal growth. Figure 5(b) shows the extent to which the Markov model dynamics is able to propagate the 24 hour forecast height anomaly. The deterministic part of the signal is not substantially larger than the random component. The covariance of the residual $\left\langle bb^T \right\rangle$ is computed in the full-space and is larger than the truncated part of the anomaly covariances.

Figure 6 shows the leading right and left singular vectors of $\mathbf{M}$. Their structure is related to the entrance of fronts and cyclogenesis in the Atlantic; 9 fronts passed through the region during the period. Using the singular values of the Markov model $\mathbf{M}$ in (9) and (10) gives as lower bound for the Talagrand ratio $\tau_{\text{var}} \geq 0.37$ and $\tau_{\text{var}} \geq 0.76$ respectively; (13) gives $\tau_{\text{vol}} \geq 0.70$ where we have conservatively used only 3 modes to compute determinants. Using more modes in the calculation of the Markov dynamics increases the singular values and decreases the lower bound as shown in Fig. 7. Of the three lower bounds, (9) is the most sensitive to the number of modes retained since it depends only on the size of the largest singular value $\sigma_1(\mathbf{M})$ of $\mathbf{M}$.

The lower bounds for τ come from assuming that the 24-hour errors project favorably onto the growing modes of the dynamics. However, the dimension of both the subspace of dominant forecast errors and of the subspace of dynamically growing modes, around 10, is small compared to the dimension of the full space 36,990. The likelihood that two arbitrarily chosen subspaces intersect is therefore small. We first compare the subspaces spanned by the forecast errors and by

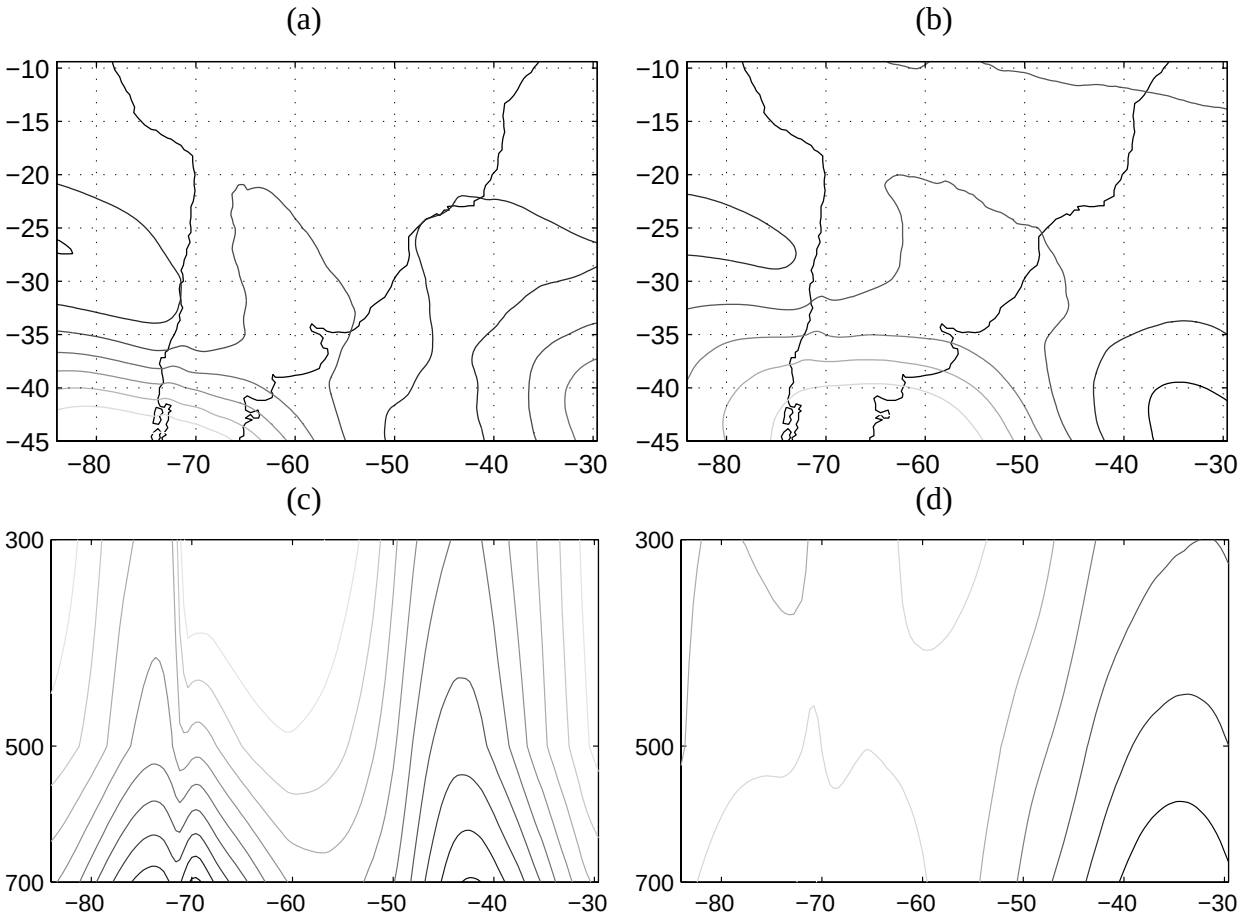

Figure 6: Contour plots of the leading right singular vector at (a) 500 mb and (c) latitude = -40.2 and of the leading left singular vector at (b) 500 mb and (d) latitude = -40.2.

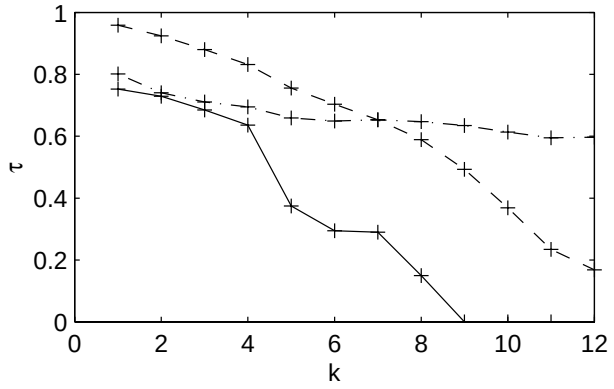

Figure 7: The lower bounds in Eq. (9) (solid line), Eq. (10) (dashed line) and Eq. (13) (dotted dashed line) as function of the number k of modes used in the Markov model calculation.

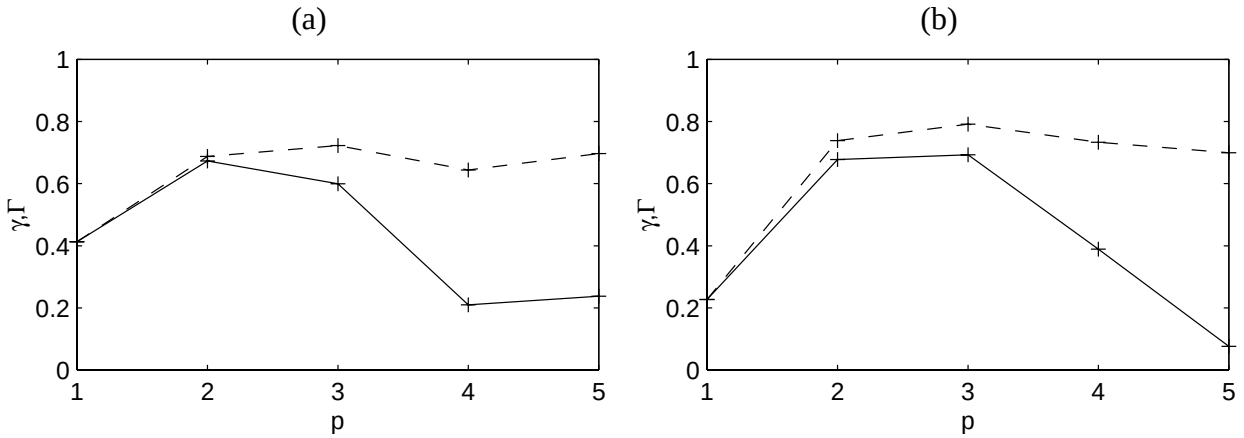

Figure 8: Plots of γ (solid line) and Γ (dashed line) (defined in the Appendix) shown as a function of the dimension p. The subspaces spanned by the (a) 24 and (b) 48 hour forecast errors and forecast anomaly fields are compared.

the forecast anomalies by comparing their principle angles (see Appendix). Figure 8 shows that there is reasonable correspondence between the subspaces spanned by the forecast error and the forecast anomaly. The correspondence is less at 48-hours than at 24 hours. At both lead times there are principle angles of about 80° indicating that there are forecast errors that project very weakly onto the subspace spanned by the forecast anomaly.

For the forecast anomaly dynamics to be able to amplify efficiently 24-hour forecast errors into 48-hour forecast errors, there must be a favorable relationship between the leading subspaces of the error covariance and the singular vectors of the dynamics. Namely, the 24-hour forecast errors must project onto the leading right singular vectors of the dynamics and the 48-hour forecast errors must project onto the left singular vectors of the dynamics. In Fig. 9 we compare the subspaces spanned by the errors and by the singular vectors of the Markov model. In Fig. 9(a) we observed that the leading modes of the 24-hour forecast errors do not project onto the leading right singular vectors of the forecast-anomaly Markov dynamics and hence are not amplified. Likewise in Fig. 9(b) the 48-hour forecast error do not project well onto the left singular vectors, suggesting that the dominant 48-hour forecast errors are not related to this dynamics.

The calculations of the subspaces associated with forecast anomaly and forecast error suggest that the forecast anomaly dynamics are insufficient alone to evolve 24-hour forecast errors in 48-hour forecast errors. This indication is confirmed in Fig. 10 where the eigenvalues of the 24 and 48 hour forecast errors are shown along with those of the propagated error covariance $\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T$. The Markov model dynamics are not able to amplify 24-hour forecast errors since the forecast error

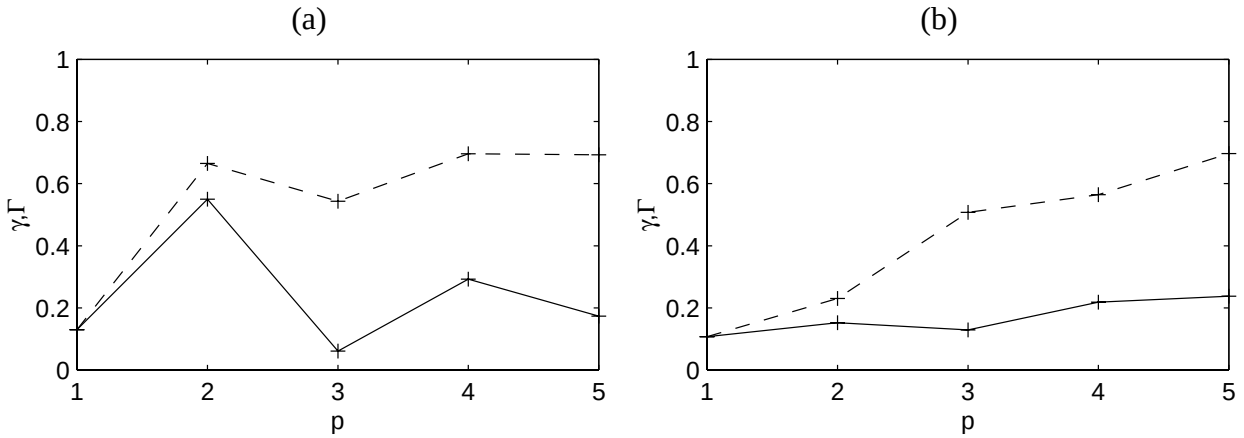

Figure 9: As in Fig. 8 but for the subspaces spanned by the (a) 24-hour forecast errors and the right singular vectors and by the (b) 48-hour forecast errors and left singular vectors are compared.

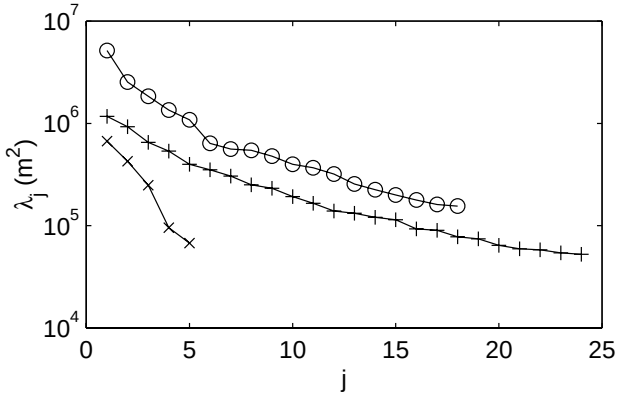

Figure 10: As in Fig. 2 with the addition of the eigenvalues of $\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T$ (solid line with x's).

are not amplified by the growing modes of the forecast anomaly dynamics. Comparing Fig. 1 with Figure 11 where the standard deviation of $\mathbf{Q}$ and $\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T$ are shown demonstrates that the model error dominates the propagated forecast error and is little different from the 48-hour forecast error.

4 Concluding Remarks

Characterization of prediction model error is important for understanding the performance of data assimilation systems. Advanced sequential data assimilation methods capable of calculating the propagated analysis error require specification of the model error. Operational 4D-Var adjoint methods assume that the model error is negligible.

A measure of the relative size of the model error is the Talagrand ratio τ, the fraction of the

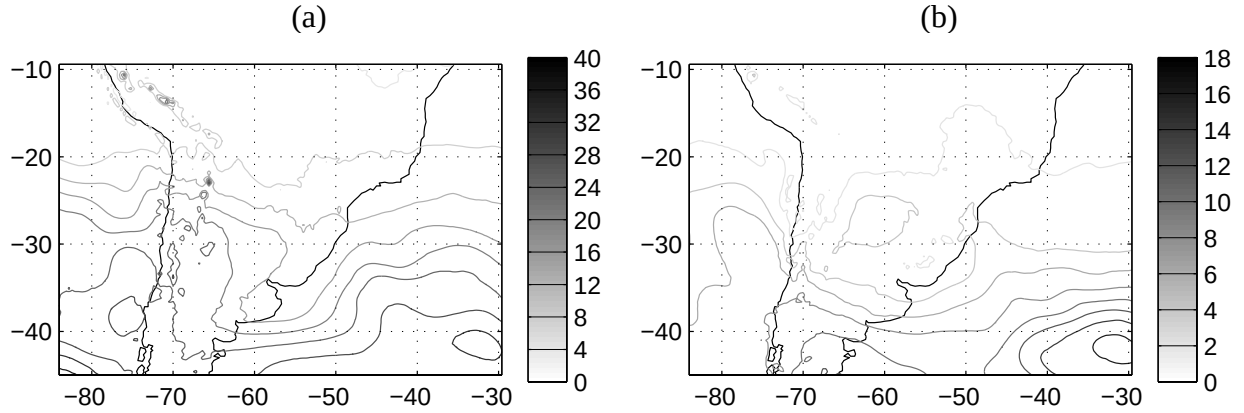

Figure 11: Contour plots of the standard deviation of (a) $\mathbf{Q}$ and (b) $\mathbf{M}\mathbf{P}_{24}\mathbf{M}^T$ shown at 500 mb. Units are meters.

forecast error due to model error. We have proposed two definitions for τ. The first is simply the ratio of the model error variance to forecast error variance; the underlying norm is the RMS one. The second is related to the volume in phase-space where errors are likely to be found. This second definition is invariant under nonsingular linear transformations; it does not depend on the choice of norm. Given the forecast error at two lead-times, the model error can be calculated. In particular, lower bounds for the Talagrand ratio can be calculated given only limited information about the dynamics.

Using data from the CPTEC Eta model and approximating the forecast error by the difference of forecast and analysis we calculated the 24-hour and 48-hour forecast height error statistics. From that information we calculated lower bounds for τ that depend on the amplification properties of the dynamics. As a linear proxy for the model dynamics we calculated a Markov model for the forecast anomaly evolution. The growth rates of this model suggest that τ is not negligible. A detailed comparison of the Markov model and the forecast errors show that they project on different subspaces and that the Markov model is not able to produce significant growth of forecast errors.

The results come with a number of caveats. First, using the analysis as a proxy for the true state when calculating forecast error is a significant assumption. Also, the NCEP analysis presents systematic and persistent differences with Eta model forecast. It would be better to use as a verifying field one whose error characteristic are better known and are stationary, for instance, observations. Second, our estimates for the model error suffer from an inconsistency between the calculated forecast error covariance and the linear dynamics. The Markov model dynamics are a poor approximation of the anomaly evolution. Perhaps a better model could be obtained by using a larger

set of variables or smaller time-steps. These alternative approaches will be investigated in future work.

Appendix

If $\mathbf{Z}_1 \in \mathbb{R}^{n\times p}$ and $\mathbf{Z}_2 \in \mathbb{R}^{n\times p}$ are orthogonal bases for two different subspaces, then there are p principle angles θ_i between the subspaces $0 \leq \theta_i \leq \pi/2$, $1 \leq i \leq p$ (Golub & Van Loan 1996, p. 603). The cosine γ of the largest principle angle is

$$\gamma = \min_i \{\cos\theta_i\} = \sigma_p\left(\mathbf{Z}_1^T\mathbf{Z}_2\right) \ , \tag{16}$$

and the average Γ of the cosines of the principle angles is

$$\Gamma = \frac{1}{p}\sum_{i=1}^{p}\cos\theta_i = \frac{1}{p}\sum_{i=1}^{p}\sigma_i\left(\mathbf{Z}_1^T\mathbf{Z}_2\right) \ . \tag{17}$$

References

Bennett, A. F. (1992). *Inverse Methods in Physical Oceanography*. Cambridge University Press.

Black, T. L. (1994). The new NMC mesoscale eta model: Description and forecast examples, *Wea. Forecasting* **9**, 265–278.

Chou, S. C. (1996). Model Regional Eta, *Climanálise* .

Cohn, S. E. (1997). An Introduction to Estimation Theory, *J. Meteor. Soc. Japan* **75**, 257–288.

Dee, D. P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Wea. Rev.* **123**, 1128–1145.

Ehrendorfer (1994). The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I Theory, *Mon. Wea. Rev.* **122**, 703–713.

Ehrendorfer, M. & Tribbia, J. (1997). Optimal Prediction of Forecast Error Covariances through Singular Vectors, *J. Atmos. Sci.* **54**, 286–313.

Epstein, E. S. (1969). Stochastic Dynamic Prediction, *Tellus* **21**, 739–759.

Farrell, B. F. (1990). Small Error Dynamics and the Predictability of Atmospheric Flows, *J. Atmos. Sci.* **47**, 2409–2416.

Ghil, M. & Malanotte-Rizzoli, P. (1991), *Advances in Geophysics*, Vol. 33, Academic Press, chapter Data assimilation in meteorology and oceanography, pp. 141–266.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*. Third edn, The Johns Hopkins University Press, Baltimore. 694 pp.

Gustafsson, N., Källén, E. & Thorsteinsson, S. (1998). Sensitivity of forecast errors to initial and lateral boundary conditions, *Tellus* **50A**, 167–185.

Houtekamer, P. L. & Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.* **126**, 796–811.

Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts, *Mon. Wea. Rev.* **102**, 409–418.

Miller, R. N., Ghil, M. & Gauthiez, F. (1994). Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems, *J. Atmos. Sci.* **51**, 1037–1056.

Schneider, T. & Griffies, S. (1999). A Conceptual Framework for Predictability Studies, *J. Climate* , in press.

Tippett, M. K. (1999). Transient moist baroclinic instability, *Tellus* **51A**, 273–288.

Yang, R., Fennessy, M. J. & Shukla, J. (1994). The Influence of Initial Soil Wetness on Medium-Range Surface Weather Forecasts, *Mon. Wea. Rev.* **122**, 471–485.