Assimilation of Standard and Targeted Observations within the Unstable Subspace of the Observation–Analysis–Forecast Cycle System

ANNA TREVISAN
CNR-ISAC, Bologna, Italy

FRANCESCO UBOLDI
LEGOS/SHOM, Toulouse, France

(Manuscript received 17 June 2002, in final form 9 July 2003)

ABSTRACT

In this paper it is shown that the flow-dependent instabilities that develop within an observation–analysis–forecast (OAF) cycle and that are responsible for the background error can be exploited in a very simple way to assimilate observations. The basic idea is that, in order to minimize the analysis and forecast errors, the analysis increment must be confined to the unstable subspace of the OAF cycle solution. The analysis solution here formally coincides with that of the classical three-dimensional variational solution with the background error covariance matrix estimated in the unstable subspace.

The unstable directions of the OAF system solution are obtained by breeding initially random perturbations of the analysis but letting the perturbed trajectories undergo the same process as the control solution, including assimilation of all the available observations. The unstable vectors are then used both to target observations and for the assimilation design.

The approach is demonstrated in an idealized environment using a simple model, simulated standard observations over land with a single adaptive observation over the ocean. In the application a simplified form is adopted of the analysis solution and a single unstable vector at each analysis time whose amplitude is determined by means of the adaptive observation. The remarkable reduction of the analysis and forecast error obtained by this simple method suggests that only a few accurately placed observations are sufficient to control the local instabilities that take place along the cycle.

The stability of the system, with or without forcing by observations, is studied and the growth rate of the leading instability of the different control solutions is estimated. Whereas the model has more than one positive Lyapunov exponent, the solution of the OAF scheme that includes the adaptive observation is stable. It is suggested that a negative exponent can be considered a necessary condition for the convergence of a particular OAF solution to the *truth*, and that the estimate of the degree of stability of the control trajectory can be used as a simple criterion to evaluate the efficiency of data assimilation and observation strategies.

The present findings are in line with previous quantative observability results with more realistic models and with recent studies that indicate a local low dimensionality of the unstable subspace.

1. Introduction

Accurate representation of the initial state is essential for the accuracy of a forecast. Analysis procedures require the combined use of observations and of a background field, usually a short-range forecast (for a review and notational conventions, see Ide et al. 1997). Observations and background are both affected by errors of different nature and characteristics. Due to flow-dependent instabilities, growth of uncertainties in the initial condition makes errors in the forecast highly variable in space and time.

In what is called the observation–analysis–forecast (OAF) cycle, a forecast relies on the analysis and the next analysis on the background field, that is, on the forecast based on the previous analysis. This procedure is particularly delicate in data-sparse areas, such as the ocean, where forecast errors that affect the analysis, not being controlled by observations, continue to grow from one cycle to the next. Because, in data-void regions, only the propagation of information from areas where observations are taken can prevent errors from growing to their saturation limit, the error in an analysis–forecast cycle is expected to depend on the structure, growth, and propagation properties of flow instabilities, as well as on the temporal and geographical distribution of observations.

The need to take supplementary observations over the ocean in order to improve the analysis and forecast

Corresponding author address: Dr. Anna Trevisan, CNR-ISAC, Via Gobetti 101, 40129 Bologna, Italy.
E-mail: a.trevisan@isac.cnr.it

in particular areas and in critical weather situations has motivated a number of studies. Strategies for selecting the location of the additional, often referred to as ''adaptive'' or ''targeted,'' observations (e.g., Snyder 1996; Palmer et al. 1998) were developed in connection with the Fronts and Atlantic Storm-Track Experiment (FASTEX) and the North Pacific Experiment (NORPEX). The general idea is that the forecast can be improved through the addition of special observational data in a particular upstream area where analysis errors are both likely to occur and likely to be fast growing (Snyder 1996). In fact, several techniques were proposed, such as adjoint-based sensitivity and singular vector calculations (Gelaro et al. 1999; Palmer et al. 1998; Bergot et al. 1999; Langland et al. 1999), the quasi-inverse linear method (Pu et al. 1997; Pu and Kalnay 1999), and an ensemble transform technique (Bishop and Toth 1999; Szunyog et al. 1999), with the aim of objectively determining where to locate the adaptive observations. In a study with a low-order system, Lorenz and Emanuel (1998, hereafter referred to as LE) were able to compare the performance of various methods on a statistical basis. They found that, at least in their simple model, ensemble-based strategies that target areas where a maximum background error is expected, are more efficient than those that target areas to which a prechosen verification region is most sensitive.

With the special observations taken during the field phases of FASTEX and NORPEX, the various methods proposed for targeting were tested relative to the limited number of real cases available from the experiments. The analysis and forecast effects of the targeted observations were positive or mixed in most cases, but, at first surprisingly, in some cases the additional data degraded the forecast (Szunyog et al. 1999; Montani et al. 1999; Langland et al. 1999). Evidence that in certain circumstances addition of the extra information provided by adaptive observations may deteriorate the forecast was also found in the experiments with a quasigeostrophic model (Morss et al. 2001) that were the continuation and extension of the results of LE. One of the motivations of our study is to address this apparent paradox. We will work on the assumption that correcting the analysis with additional data is not necessarily beneficial to the forecast, unless the projection of the error on the unstable subspace is reduced.

The focus of the paper is on building a method of targeting and assimilating observations that is consistent with the dynamical evolution of the errors: ideally the analysis should satisfy the governing equations of the model atmosphere, and it should coincide with the state that the true trajectory goes through at the analysis time.

An important consideration is that the analysis error that is going to grow in the next forecast is, to a large extent, due to the same flow instabilities that have generated the background error. The aim of ensemble perturbations systems that are part of the forecast practice is to sample the initial error probability density function so as to capture the structure of growing errors. The importance of the flow-dependent instabilities was recognized by Kalnay and Toth (1994) who used bred vectors to reduce the *errors of the day* in the analysis cycle. Using time-dependent statistics to obtain a reliable estimate of the background error covariance matrix is at the basis of the recently developed techniques known as ensemble Kalman filters (EKFs; Evensen and van Leewuen 1996; Hamill and Snyder 2000; Houtekamer and Mitchell 2001).

The rapid progress in the operational implementation of advanced data assimilation is founded upon the theoretical development of variational- and sequential-estimation approaches. The connection between optimal interpolation and 3D variational assimilation (3DVAR), and Kalman filter and 4D variational assimilation (4DVAR) are now well established (Ghil 1989; Ide et al. 1997; Daley 1991; Kalnay 2002; Bennett 2002; Uboldi and Kamachi 2000).

Basic questions to be addressed through data assimilation and observation system simulation experiments (OSSE) concern

1) the observability of the system, that is, how many observations are necessary in a given time interval to determine its state;
2) the propagation of information from data-dense to data-sparse regions.

Precise statements of the problems can be found in Ghil and Malanotte-Rizzoli (1991), and a wealth of quantitative observability results is reviewed in Ghil (1997). To mention just a few, Todling and Ghil (1994) and Ghil and Todling (1996) have shown for a barotropic and baroclinic, linear 2D model that one observation per known unstable pattern is sufficient to track the solution, but the need for a very small number of observations has also been demonstrated in strongly nonlinear models. Another important finding (Patil et al. 2001) is that ensemble perturbations frequently evolve into locally low-dimensional subspaces.

Issues related to propagation of information, in particular how data density and advection influence the estimation error covariance, were covered by Ghil et al. (1981).

The basic idea behind the method that we propose consists in making corrections to the background that are confined to the unstable subspace, so that the analysis increment has the same spatial structure as the dominant instabilities of the system. Because the OAF cycle is a system forced by observations, we will construct ensemble perturbations consistent with the equations describing its stability. Once the unstable vectors, which we refer to as *assimilating* vectors, have been identified, their amplitude is obtained by minimizing the distance of the analysis solution from the observations.

In an adaptive observations context, this involves two steps.

1) Find the geographical area where observations are needed in order to reduce the actual analysis error, that is, errors that could grow from uncertainties present in the previous analysis. To this end, ensemble perturbations are used to track the instabilities that are growing at a particular time in a particular area along the OAF cycle.
2) Use the data from standard and additional observations in combination with the dynamical, rather than statistical, information on the structure and growth of background errors to reduce the analysis error and, at the same time, to control the instabilities that are going to subsequently affect the forecast.

In this paper we will develop the formalism in the general context of an arbitrary number of observations and ensemble perturbations. The 3DVAR cost function is minimized, but the analysis increment is confined to the unstable subspace of the OAF cycle solution. The estimate of the background error covariance is obtained on a dynamical rather than statistical basis.

The procedure is then implemented in the simple context of the 40-variable model of LE, suitable for the statistical evaluation of the comparative performance of adaptive observation strategies. For comparison with previous results, we make use of a simplified analysis solution with a single adaptive observation and a single assimilating vector at each analysis time. The merits of the present targeting-assimilation method are demonstrated by the remarkable improvement to the analysis and forecast that is achieved.

We then discuss the stability of control solutions with and without forcing by standard and adaptive observations. We argue that a negative growth rate can be considered a necessary condition for convergence to the *truth.* The leading exponent of different OAF schemes is computed in the LE model and suggested as a criterion to estimate the relative efficiency of assimilation methods and adaptive observation networks.

2. The Lorenz and Emanuel model and targeting strategies: Previous results

The model is a low-order chaotic system (Lorenz 1996), introduced by LE for the study in question and exploited by several authors for the same purpose. Its variables $x(j)$, $j = 1, J$, represent the values of a meteorological quantity at equally spaced geographic sites along a latitudinal circle. The evolution is governed by

$$\frac{d}{dt}x_j = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F, \tag{1}$$

with periodic conditions on the domain.

The nonlinear quadratic terms in (1) simulate advection, the linear term represents dissipation, and F is a constant external forcing. The time step used by LE is equal to 0.05 and corresponds to 6 h if dissipative decay time in model units is interpreted to be 5 days. The number of points in the cyclic grid is 40. A trajectory obtained by integrating the system equations with forcing $F = 8.0$ is considered to be the true evolution of the physical system. A systematic model error is assumed to affect the forecast and is introduced by integrating the model equations with a slightly different forcing, $F' = 7.6$.

The experiments designed by LE to evaluate the relative merits of the various targeting strategies consist of idealized observation system simulations. Standard observations are located over land (grid points 21–40) at analysis times (every 6 h). One supplementary observation is located over the ocean, at one of the grid points from 1 to 20. All observed values are obtained by adding an observational error to the *true* value; the observational error is normally distributed with zero mean and standard deviation $\sigma = F/40 = 0.2$. The analysis is obtained from the first guess by replacing the background values with the observed values at the sites of the standard and of the single adaptive observations.

From the numerous tests performed by LE, three main targeting strategies turned out to give the best results: multiple breeding (LE-MB), singular vectors (LE-SV), and multiple replication (LE-MR), with the last one outperforming all the other methods. These strategies are variants of methods used for generating perturbations in ensemble forecasting systems that are discussed at length in the meteorological literature. The trajectory that is perturbed is the control, that is, the result of the complete OAF procedure. The site to be targeted is the one where maximum background error is estimated to occur based on the difference between the control and the perturbed trajectories of the ensemble.

In LE-MB each perturbation vector was renormalized every 6 h by setting the sum of the squares of its components equal to the initial value. LE-MR is a modification of LE-MB, in which an additional error (with the same distribution as the observational error) was introduced in the observations to be assimilated in the perturbed trajectories. In LE-MB and in LE-MR the targeted site is the one where the sum, over an ensemble of 15 perturbations, of the squares of the perturbation components attains its maximum value. In LE-SV the targeted site coincides with the largest component of the most rapidly growing perturbation introduced 10 days earlier. Lorenz and Emanuel (1998) adopted final, rather than initial, singular vectors and chose 10 days, which might be considered a rather long optimization time, because these choices were found to produce the best results.

The supplementary observation was assimilated in all, perturbed and unperturbed, analysis. This is an important feature common to all strategies tested by LE that we will retain in our formulation.

Because of the choices made in the design of the experiments (in particular the optimization time and selection of the final leading SV), a certain similarity between LE-MB and LE-SV could be expected. For fur-

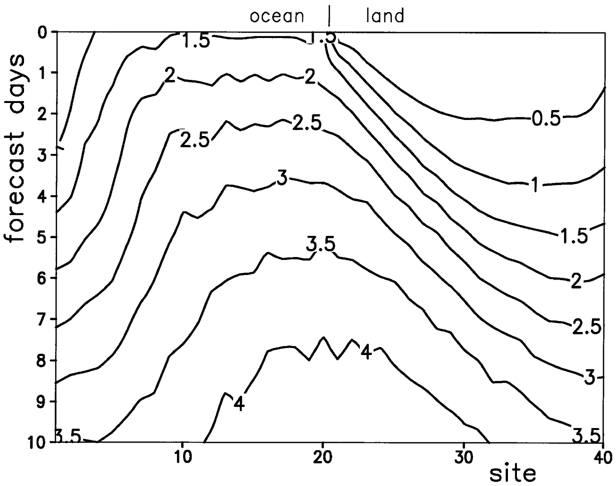

FIG. 1. Twenty-year-average analysis and forecast error from the multiple replication experiment with one adaptive observation of LE as a function of the spatial coordinate and forecast range. Land, with standard observations, is on the eastern half of the domain; the western side is the ocean, where a single targeted observation is located. The analysis error over land is, by definition, equal to the rms observational error (=0.2) and gradually increases to reach, at ocean grid point 9, a value of about 1.5, which remains almost constant within the eastern half of the ocean domain (grid points 9–20). The frequency of observation as a function of the site follows a pattern very similar to that of the analysis error, with approximately two-thirds of the adaptive observations being uniformly distributed over the eastern half of the ocean domain.

ther details on the implementation of these methods, we refer the reader to the original paper.

The results of LE-MB and LE-SV are similar, but very slightly better for LE-SV. A more significant improvement is obtained by LE-MR. Figure 1 reproduces the results of LE-MR obtained from a 20-yr rerun of their experiment. The average analysis and 10-day forecast errors are shown for each (ocean and land) site. Lorenz and Emanuel (1998) interpret the success of multiple replication as being due to its ability to diversify the perturbations, providing a richer variety of error structures with the same likelihood of representing the true error.

Other authors have used the LE model to test targeting strategies. Berliner et al. (1999) adopted a statistical design to find the optimal location on the basis of the esimated forecast error covariance matrix at a desired future time.

Hansen and Smith (2000) applied a 1024-member ensemble Kalman filter assimilation scheme, combined with MB and SV and obtained a remarkable reduction of the analysis and forecast error. An important conclusion of their work is that, in the limit as the analysis error approaches zero, the information about future instabilities becomes most valuable. In fact, when the observational error is reduced (by a factor of 16 with respect to LE), the adaptive observation strategy based on *future* singular vectors was shown to outperform multiple breeding, with bred vectors computed in the *past.* It should be noted also that these authors used a scalar product based on the inverse of the analysis error covariance matrix, as estimated by the ensemble technique.

3. Formulation of the approach to targeting and assimilation of adaptive observations

The building blocks of our approach are the following.

1) As mentioned in the introduction, other authors have shown that the assimilation of additional observations is not necessarily beneficial to the forecast and can, in some cases, lead to an increase of the forecast error. Usually, only some components of the state vector are observed: in the simplest case these components are represented by meteorological variables at particular locations where observations are taken. Considering the simple example of an assimilation scheme that replaces background values by observed values at observational sites, one can easily see that, although the analysis error is reduced, the same may not be true for the projection of the analysis error on the most unstable vectors. In such a case, a reduction of the forecast error cannot be expected. Therefore, our first goal is to estimate and reduce the projection of the actual analysis error onto the most rapidly growing perturbations.
2) A second point is in regards to the nature of the error. The errors in the background that one wishes to reduce by introducing adaptive observations are errors that were free to grow in the OAF cycle due to the lack of data; in fact, at sites where regular observations are taken, errors can grow only during the time span between analysis. In data-void areas, instead, the unstable structures are free to grow for longer time periods during the OAF cycle. Errors that were present at some stage in a particular control trajectory (the result of the OAF procedure), after some sufficiently long time during which many analysis and forecast steps are performed, always converge to some typical structures that depend both on the particular analysis and forecast process and on the flow-dependent instabilities. ‘‘After the growing modes have grown and the decaying modes have decayed’’ the background error ‘‘has approximately the structure of a combination of the more rapidly growing modes rather than a combination of all modes’’ (LE).

If the analysis increment is a linear combination of these structures with amplitude coefficients determined by a fit to the observations, the correction to the analysis will be consistent with the dynamical evolution of the actual error: along the unstable directions, the attractor is continuous and one can move from one state to another state that also belongs to the attractor and lies on a nearby trajectory (Lo-

renz 1984). The correction will not only be consistent, but will also minimize subsequent error growth; in fact, super-Lyapunov growth has been shown to be associated with transient effects due to the nonorthogonality of locally stable and unstable vectors (Lacarra and Talagrand 1988; Trevisan and Pancotti 1998).
3) The third point concerns the identification of the unstable subspace of the OAF system, which is forced by observations, and to this end the perturbed trajectories must also be subject to the same forcing, a feature present in LE and in ensemble Kalman filters.
4) A final important point concerns observability. This property is strictly connected both to the stability of the solution and to the flow of information throughout the domain. The number and frequency of observations necessary to counteract the instabilities are proportionally related to their growth rate and to the dimensionality of the unstable subspace. The distribution and specific positioning of observations with respect to the growing error structures play an important role in the overall picture.

4. Analysis solution: Confinement in the unstable subspace

Assume that, by proper ensemble techniques, we are able to find N independent vectors that describe the unstable subspace or at least the leading instabilities of the OAF system. The background error has an important component in this subspace and can be expressed as a linear combination of the unstable vectors, referred to as assimilating vectors. By assimilating a sufficiently large number M of observations, $M \geq N$, we can estimate the component of the error that projects on the unstable subspace and eliminate it from the analysis.

Let the set of orthonormalized unstable vectors be given by $\mathbf{e}_k$, $k = 1, N$, and let $\mathbf{y}^o$ be the M-dimensional observation vector. The analysis increment confined to the unstable subspace spanned by the $\mathbf{e}_k$ is given by

$$\delta\mathbf{x}^a = \mathbf{x}^a - \mathbf{x}^b = \sum_{k=1}^{N} a_k\mathbf{e}_k \equiv \mathsf{E}\mathbf{a}, \tag{2}$$

where $\mathbf{x}^a$ is the analysis, $\mathbf{x}^b$ the background state, E is the $I \times N$ matrix whose columns are the $\mathbf{e}_k$ vectors, and $\mathbf{a}$ is the vector whose components are the coefficients a_k to be determined.

In view of (2), the cost function for an incremental (linearized) 3DVAR analysis reads

$$J = (\mathsf{H}\mathbf{x}^b + \mathsf{HE}\mathbf{a} - \mathbf{y}^o)^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{H}\mathbf{x}^b + \mathsf{HE}\mathbf{a} - \mathbf{y}^o) + \mathbf{a}^{\mathrm{T}}\mathsf{E}^{\mathrm{T}}\mathsf{B}^{-1}\mathsf{E}\mathbf{a}, \tag{3}$$

where H is the (Jacobian of the) observation operator, and B and R are the covariance matrices of the background and observational error, respectively.

Setting $\nabla_{\mathbf{a}} J = 0$, we get

$$\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathsf{HE}\mathbf{a} + \mathsf{H}\mathbf{x}^b - \mathbf{y}^o) + \mathsf{E}^{\mathrm{T}}\mathsf{B}^{-1}\mathsf{E}\mathbf{a} = 0. \tag{4}$$

The solution is

$$\mathbf{a} = (\mathsf{E}^{\mathrm{T}}\mathsf{B}^{-1}\mathsf{E} + \mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{HE})^{-1}\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1} \times (\mathbf{y}^o - \mathsf{H}\mathbf{x}^b) \quad \text{or} \tag{5a}$$

$$\delta\mathbf{x}^a = \mathsf{E}(\mathsf{E}^{\mathrm{T}}\mathsf{B}^{-1}\mathsf{E})^{-1}\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}[\mathsf{HE}(\mathsf{E}^{\mathrm{T}}\mathsf{B}^{-1}\mathsf{E})^{-1}\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}} + \mathsf{R}]^{-1} \times (\mathbf{y}^o - \mathsf{H}\mathbf{x}^b). \tag{5b}$$

We will consider the two following cases.

a. *The background term in the cost function is neglected*

The solution obtained by minimizing the cost function (3) after dropping the last term is

$$\mathbf{a} = (\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}\mathsf{HE})^{-1}\mathsf{E}^{\mathrm{T}}\mathsf{H}^{\mathrm{T}}\mathsf{R}^{-1}(\mathbf{y}^o - \mathsf{H}\mathbf{x}^b). \tag{6a}$$

When $M = N$, the matrix HE is square and in relevant cases, invertible and the solution simplifies to

$$\delta\mathbf{x}^a = \mathsf{E}(\mathsf{HE})^{-1}(\mathbf{y}^o - \mathsf{H}\mathbf{x}^b), \tag{6b}$$

where the matrix R has dropped out and the analysis exactly fits the observations.

If, in particular, we consider the case of a single observation $N = M = 1$, then E consists of a single column vector $\mathbf{e}_1$ and the analysis increment is given by

$$\delta\mathbf{x}^a = \frac{y^o - x^b(\mathrm{iobs})}{e_1(\mathrm{iobs})}\mathbf{e}_1, \tag{7}$$

where the observation operator was simplified by the assumption that one scalar component of the state vector is observed. Here, $x^b(\mathrm{iobs})$ and $e_1(\mathrm{iobs})$ indicate the components of the vectors $\mathbf{x}^b$ and $\mathbf{e}_1$ at observation point iobs.

This simple case ($N = 1$) is readily interpreted. The correction to the background has the structure of the unstable perturbation and an amplitude to exactly fit the observation. In the hypothesis that the background error, at a given time, has the same structure as the assimilating vector, that is, it points in the direction of $\mathbf{e}_1$, a single exact observation would be sufficient to reduce the analysis error to zero.

b. *The background error is confined within the N-dimensional subspace*

If E is assumed to span the N-dimensional subspace with $N < I$, where the background error $\boldsymbol{\eta}^b$ is confined, that is,

$$\boldsymbol{\eta}^b = \mathsf{E}\boldsymbol{\gamma}, \tag{8a}$$

then its covariance matrix

$$\mathsf{B} = \mathsf{E}\boldsymbol{\Gamma}\mathsf{E}^{\mathrm{T}} \tag{8b}$$

is of rank N and is not invertible.

However, the N-dimensional basis of the unstable subspace $\mathbf{E}$, together with its orthogonal complement, defines a basis on the whole space. In this new basis, the background error covariance matrix can be seen as a 2×2 block matrix, where only the first (upper and left) block has nonzero elements, the (N, N) invertible matrix $\mathbf{\Gamma}$, which represents the background error covariance matrix in the N-dimensional subspace. A reduced-order cost function for this problem can be written that is consistent with the strong constraint (2) that the analysis increment is confined in the unstable subspace.

The cost function becomes

$$J = (\mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}\mathbf{a} - \mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}\mathbf{a} - \mathbf{y}^o) + \mathbf{a}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{a}, \tag{9a}$$

and the solution of the minimization problem reads

$$\delta\mathbf{x}^a = \mathbf{E}\mathbf{\Gamma}\mathbf{E}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{E}\mathbf{\Gamma}\mathbf{E}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}})^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b). \tag{9b}$$

The same solution can be obtained by imposing (8a) in the derivation of an optimal interpolation or Kalman filter analysis, and it does not depend on the particular scalar product chosen.

Expression (8b) can be seen as a (suboptimal) substitute of the Riccati equation for covariance evolution in Kalman filters. In fact, (9b) formally coincides with the analysis solution used in ensemble Kalman filter approaches based on reduced-order background matrices. For an extensive discussion and review of reduced-order filter techniques see Kalnay (2002).

Because we have identified the basis vectors $\mathbf{E}$ with the (leading) unstable directions, the present formulation suggests a simplified approach to reduced-order Kalman filter. In contrast with ensemble Kalman filters, where the background error covariance matrix is estimated from the dispersion of the ensemble members, in our approach we assume to know the unstable directions of the system. The order reduction is obtained by estimating only a (small) number N of unstable vectors. Given the local (leading) exponents, the analysis error projection onto the corresponding unstable directions amplifies accordingly and this completely determines the background error covariance matrix in the (reduced) subspace (8b).

In particular, when $N = 1$, (8b) reduces to

$$\mathbf{B} = \gamma_1^2\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}, \tag{10}$$

where γ_1 is the expected component of the background error in the direction of $\mathbf{e}_1$.

With this prescription, $M = 1$ and $\mathbf{R} = \sigma^2\mathbf{I}$, (9b) reads

$$\delta\mathbf{x}^a = \frac{e_1(\mathrm{iobs})[y^o(\mathrm{iobs}) - x^b(\mathrm{iobs})]}{[e_1(\mathrm{iobs})^2 + \sigma^2\gamma_1^{-2}]}\mathbf{e}_1 \tag{11}$$

to be compared with (7). Kalnay and Toth (1994) used a similar expression to account for the background errors in the direction of bred vectors.

In the application to the 40-variable model, we will make use of (7) rather than (9b) or (11) to directly compare our results with those of LE. Use of (9b) or (11) would in fact require an estimate of $\mathbf{\Gamma}$ or γ_1 in addition to the associated vectors that can be easily obtained from the evolution of analysis error in the corresponding subspace and will be the subject of future investigation.

5. Application to the LE model of the targeting–assimilation method

Focusing our attention on the problem of adaptive observations, we consider the application of the method to the LE model, where some simplifications are in order. For comparison with previous results, only one adaptive observation is made at each analysis time and a single assimilating vector is used. The observation is placed at the location where the current leading unstable structure $\mathbf{e}_1$ has amplified the most. As a consequence, the analysis increment is maximum at the observation location.

The construction of perturbations aimed to capture the unstable directions to be used in the assimilation is an important and delicate point. As further discussed in section 6 and appendix B, which deal with its stability, in the OAF system, forced by observations, the growth of errors and the dimension of the unstable subspace are reduced. For consistency, the perturbed trajectories must also be forced by observations.

The simple and efficient procedure used to identify the unstable directions is similar to breeding, except that all perturbed trajectories are subject to the same OAF cycle as the control, as in LE. As such, the perturbations, whose growth is constrained by the observation–assimilation process, will take up the structure of the instabilities of the OAF solution.

In the present application, a perturbation is introduced every 6 h by adding to the control analysis a random error at each grid point, and is grown for 10 days (40 observation times) before being used. The perturbation, after being used, is discarded and a new one is introduced.

The adaptive observation is located at the site where the *current* perturbation vector has its largest component. The current perturbation, that is, the perturbation inserted in the analysis 10 days earlier, is used to estimate the present most unstable vector $\mathbf{e}_1$ (the assimilating vector) and to assimilate the observation by applying (7) to the control trajectory and to the remaining 39 perturbed trajectories.

Because $M = N = 1$, the unperturbed and the perturbed analysis are all equal to the observed value at the targeted location (all perturbations have a zero component at the observation site). At the same time, (7) implies that the analysis increment at surrounding points is nonzero, but has the same spatial structure as the current perturbation: if the observation was perfect, er-

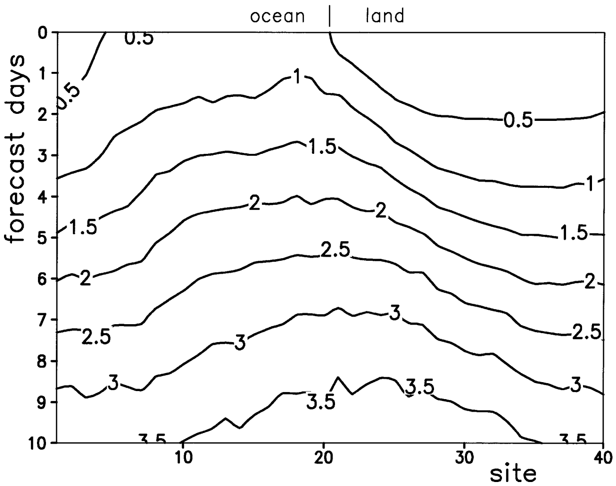

FIG. 2. Same as Fig. 1, but using the observation–assimilation method presented in the text. The analysis error grows from 0.2 to 0.6 between grid points 1 and 5; and is approximately constant throughout the ocean domain, except for the easternmost grid points (15–20), where it reaches a maximum of 0.8. Following the spatial pattern of the analysis error quite closely, the frequency of observation is almost constant throughout most of the ocean domain, except for a few grid points near the east and west coast, where we find the smallest and the largest number of observations, respectively.

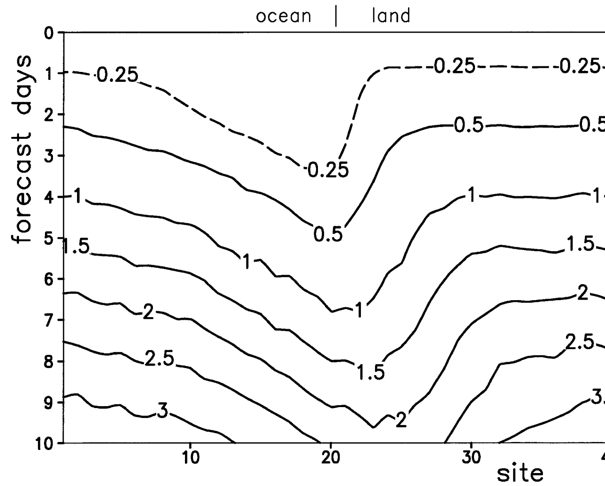

FIG. 3. Same as Fig. 2, but without model error ($F = 8$) and with a perfect adaptive observation. The analysis error over the ocean is everywhere smaller than over land, being on average less than 0.1. The frequency of observations is similar to that of the expt of Fig. 2.

rors in the direction of $\mathbf{e}_1$ would be completely eliminated from all, perturbed and unperturbed, trajectories.

It is worth at this point to recall that the main difference between our experiment and those of LE is in the use of (7) to assimilate targeted observations. We have conducted several additional experiments whose results will not be shown. First, we varied the breeding time from 1 to 20 days; provided the breeding time is at least 3 days, the results are not very sensitive to this parameter. We also implemented the different perturbation methods used by LE and referred to in section 2 as LE-MB, LE-SV, and LE-MR. The differences in the results among the various experiments are only minor if compared with the substantial improvement obtained using any perturbation strategy in combination with (7).

Figure 2 shows results analogous to those of Fig. 1, but obtained with the procedure described above and making use of (7). The improvement in the ocean area is very significant: comparing Fig. 2 with Fig. 1, it is not until 3 days into the forecast that we find, in the statistics, errors as large as those present in the analysis. In the present experiment, a single observation over the ocean is sufficient to reduce both the analysis and forecast error to such an extent to make it comparable to the error over land. The success of the experiment is particularly encouraging in view of the fact that we have used a single observation and a single assimilating vector, and one can expect that the error will be further reduced if a larger number of vectors and observations is used. On the other hand, if assimilations are made at relatively short time intervals, the various unstable perturbations that, at a certain time and location, contribute to growth of the background error are continuosly being reduced. The number of vectors taken into account and the time interval between corrections can in fact be considered mutually compensating factors.

We now turn to the question of whether the number of the unstable directions and observations are sufficient for the system to be observable, having in mind that this property depends upon various factors, such as the instability growth rate, the dimension of the unstable subspace, and the frequency of observations. In order to investigate this point we performed an experiment under perfect model conditions and a single perfect adaptive observation over the ocean. In this experiment, apart from the initial condition, the imperfect observations over land constitute the only source of error.

The results of Fig. 3 show a major improvement of the analysis and forecast over the ocean and demonstrate the possibility of controlling local unstable growth in this model by means of the single observation over the data-void region. Now it is the correct information from the simple perfect ocean observation that propagates westward, leading to a significant reduction of the error over land throughout the entire forecast range.

It needs to be mentioned that in the present model it was necessary to confine the region of influence of (7) by applying a masking function to the perturbation $\mathbf{e}_1$ in order to avoid an instability that appeared in the solution under particular circumstances. We refer to appendix A for details on this point.

6. Estimating the optimality of observation–assimilation schemes from the stability of the solution

In the present section we study the stability of the solution of observation–assimilation schemes. We expect that the ability of a given scheme to control error growth and force the solution to remain close to the true trajectory should be reflected in a reduced divergence rate.

The truth is represented by a solution of the dynamic equations governing our model atmosphere; this solution is unstable with respect to infinitesimal perturbations of the initial condition. A particular observation–analysis scheme provides us with another solution, referred to as the control, which is the outcome of the complete analysis cycle. Integration of the original dynamic equations between successive analysis times provides the background, but the analysis solution is then ‘‘forced’’ to remain close to the observations. In the scheme that we have proposed, the forcing toward the observations is done in a way consistent with the error dynamic evolution, so that the corrected analysis can be viewed as a state that belongs to a nearby trajectory of the system.

Thus, the control solution can be regarded as the solution of a set of forced equations that is more complex than the original model. The solution of such equations may itself be sensitive to initial conditions, but, as can be anticipated, its stability properties may very well differ from the stability properties of the original system. Following Ghil et al. (1981), who first introduced this distinction, and Ide et al. (1997), we outline the stability problem of a system with and without data-forcing in appendix B, which provides the framework for the following discussion.

If the control solution is stable, then starting with slightly different initial conditions, compatible with the observational error and treated exactly as the control, we will in practice end up with one and the same solution. This does not mean that the solution will converge to the truth, in view of imperfections in the model and in the analysis scheme or in the case that the error in the initial condition is not sufficiently small.

If, instead, the control solution is unstable, it will be at all times dependent on the initial condition and will represent but one of the possible solutions of the observation–analysis scheme. Even with a perfect model and perfect observations we will never know which one of the possible solutions is the truth.

We can thereby consider the stability of the control solution as a necessary, not sufficient, condition for convergence to the truth. With regard in particular to an assimilation scheme including adaptive observations, its ability to target areas where errors tend to be large and fast growing may eventually halt their growth and subsequently keep it within bounds.

The rate of growth of perturbations of the control can give us a measure of the efficiency of a given scheme to keep the error small: a scheme that renders the control stable will presumably perform better also in reducing the actual analysis error. Thus, the estimate of the growth rate of perturbations of the control provides us with a simple means of comparing the potential efficiency of different schemes and constitutes a valid alternative to the statistical evaluation of the actual analysis error. In fact, the latter is feasible only in idealized experiments in which the true evolution of the system is assumed to be known, or by means of complex parameter estimation algorithms (Dee et al. 1985; Navon 1997).

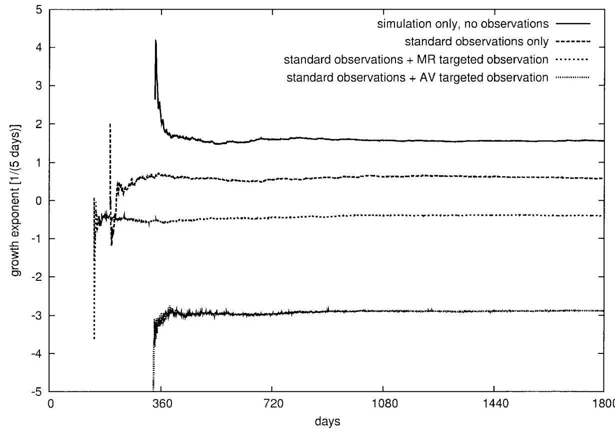

FIG. 4. Average growth rate of the leading unstable vector as a function of time for the following solutions: no observations, standard observations over land, standard observations and one adaptive observation LE-MR experiment, and standard observations and one adaptive observation with the observation–assimilation scheme presented in the text. Averaging starts at the time when a set of initially arbitrary perturbations have approached each other.

In the context of the present experiments with the LE model, by comparing different observation–analysis schemes, we will show that, indeed, a reduction of the growth rate corresponds to a reduction of the analysis error obtained with that scheme.

The rate of growth of perturbations are computed in the usual manner. A perturbed trajectory is generated by adding a small error, randomly drawn from the same distribution as the observational error to the initial condition. The perturbation is renormalized at regular time intervals so that errors evolve linearly. Thus, when we apply this procedure to the original equations and compute the average growth rate, we recover the leading Lyapunov exponent of the system. Analogously, to estimate the stability of the solution of a given OAF scheme, we add an error to the control analysis and compute the growth rate of the small, linearly evolving perturbation. In such a case, the perturbation is kept small by renormalizing its amplitude, and the perturbed trajectory is subject to the same observation–analysis procedure as the control trajectory of the given scheme. Whether only standard or adaptive observations are available, they are assimilated in the perturbed trajectory in the same way as they are assimilated in the control at each analysis time. Figure 4 shows the linear growth rate, averaged along the trajectory, of perturbations to the following solutions: the original system, the control solution with standard observations over land, the control solution with the adaptive observation multiple replication scheme of LE, and the solution with the adaptive observation–assimilation scheme that we have presented. The growth rate corresponding to the solution with standard observations over land is smaller than the Lyapunov exponent of the free system but is still positive. The addition of one adaptive observation is sufficient to render the control solutions stable both in the MR experiment of LE and in our experiment, but the rate of convergence is much faster within our scheme. We can thus regard the efficiency of a particular method in reducing the actual analysis and forecast errors observed in the statistics as a consequence of the stability of its solution.

7. Conclusions

The main features of the formulation of the data assimilation process within an OAF cycle presented in this paper are the following.

1) The correction to the background state (the analysis increment) is confined to the subspace of the dominant unstable directions of the system, that is, of the OAF routine. The set of coefficients that provide the analysis solution is determined by minimizing the 3DVAR cost function. A simplified solution is obtained in the case of a single unstable direction and a single observation.
2) The unstable directions of the OAF system solution are estimated by breeding perturbations subject to the same observation–analysis–forecast process as the control solution.
3) A distinction is made between the stability of the forecast solution and the stability of the analysis-cycle solution, the forcing imposed by the observational constraints having the effect of reducing the instability of the solution. The growth rate of the leading instability mode within a particular OAF scheme could be compared with that of the original model equations or of other schemes. It is suggested that a negative growth rate is a necessary condition for convergence to the *truth*, and that the minimization of the exponent can be considered a criterion to estimate the optimality of observation and assimilation systems.
4) The question about the observability of the system is posed in terms of the number of observations necessary to determine the state of the system, given the number of unstable structures that develop locally in space during a certain time interval.

The results of the application to the LE 40-variable model with standard observations over land and one adaptive observation over the ocean have shown the following.

1) Using our method to assimilate the adaptive observations in a 20-yr rerun of the observation system simulation experiment of LE, average analysis errors over the ocean were reduced by 50%. Only after 3 days into the forecast, they became as large as those present in the analysis. This improvement was achieved using the same amount of information and practically no additional computational cost.
2) A perfect model experiment with a perfect adaptive observation further demonstrated the success of the deterministic assimilation scheme: with a single observation over the oceanic data void region covering half of the domain it was possible to determine the state more accurately than over land. The correct information over the ocean propagates eastward with beneficial effects to the forecast over land.
3) The leading exponent of the scheme that we have implemented has a large negative value, confirming the efficiency of the proposed adaptive observation assimilation procedure in controlling error growth by stabilizing the solution.

The readers are cautioned about the limited applicability of the LE model to realistic data assimilation. However, the questions we have posed about the stability and observability of OAF systems can be addressed, by the same algorithmic approach, in more realistic models in which some of the qualitative conclusions can still be expected to hold. The reduction of the number and value of positive exponents and consequently of the number of observations necessary to control the instabilities in the data-forced OAF system opens some optimistic perspective also for the real world assimilation problem.

Acknowledgments. The insightful comments and constructive criticism of the reviewers, Michael Ghil and Jim Hansen, were extremely helpful. The authors also wish to thank Maurizio Fantini, Remy Baraille, and Son Hoang for useful comments and discussions. Work was supported by the MIUR Special Project: SINAPSI/Ecosistemi Marini.

APPENDIX A

Use of a Modulating or Masking Function to Regionalize the *Assimilating* Vector

In the application of the scheme to the LE model with a single observation and a single unstable perturbation at each analysis time, it was necessary to ''regionalize'' the *assimilating* vector correction in order to prevent uncontrolled error growth. This problem, however, occurred only in a few instances and in particular circumstances, in the 20-yr assimilation experiment that was run. In such cases, the solution diverged from the *truth*, making successive corrections based upon perturbations of the control totally erroneous. Close inspection revealed the cause of the problem to be associated with the following concurrence of events. The perturbations used in sequence to target and assimilate the observation appeared to be the result of the superposition of different modes and showed spatially separated maxima or minima; also the background error appeared to be the superposition of the same modes, but with different amplitudes, occasionally with the opposite sign. While the observation and correction in correspondence to the maximum perturbation amplitude led, as in all other cases, to a reduction of the error in that area, the presence of the secondary peak led to a correction with the wrong sign. When this event occurred several times in a row, along the cycle, the solution became unstable.

For this reason, it was chosen to regionalize the correction. The current perturbation, that is, the assimilating vector is modulated by a function that goes to zero outside an area surrounding the observation. Among the many possible choices, we decided to use a Gaussian function. After some trials, the decay scale was set equal to three grid points. With this value, the region of influence of the assimilation correction is sufficiently wide that the results, compared with those obtained rejecting the few unstable cases that were found during the assimilation, are not significantly altered by the application of the masking. It is important to note that the value chosen allows the assimilation of a single observation to affect one-third of the whole ocean area, and that correlations are active on a much larger distance than what is usually done in space analysis.

APPENDIX B

The Stability of the Data-Forced System

The time evolution of a real system is approximated by the nonlinear model equations

$$\frac{d\hat{\mathbf{x}}}{dt} = M[\hat{\mathbf{x}}], \tag{B1}$$

where $\hat{\mathbf{x}}$ indicates the estimate of the *true* state $\mathbf{x}^t$. In the sequential estimation approach, the true state is assumed to evolve according to a stochastic differential equation

$$d\mathbf{x}^t = M[\mathbf{x}^t] \cdot dt + d\boldsymbol{\eta}, \tag{B2}$$

where $\boldsymbol{\eta}$, the model error, is a Wiener process (Ide et al. 1997).

In the absence of data, the evolution of perturbations is obtained by linearizing the equations about the estimated trajectory

$$\frac{d\delta\hat{\mathbf{x}}}{dt} = \mathbf{M}[\hat{\mathbf{x}}] \cdot \delta\hat{\mathbf{x}}. \tag{B3}$$

When data are available, the assimilation process forces the estimated trajectory toward the truth.

Observations in general are incomplete and affected by errors

$$\mathbf{y}^o = H[\mathbf{x}^t] + \varepsilon^o, \tag{B4}$$

where ε^o represents the observational error, including that specific to the observation operator H.

If observations are assimilated at intermittent times t_k, the evolution of the data-forced system is given by

$$\frac{d\hat{\mathbf{x}}}{dt} = M[\hat{\mathbf{x}}(t)] + \sum_k \delta(t - t_k) \cdot \mathbf{K}_k \cdot \{\mathbf{y}_k^o - H_k[\hat{\mathbf{x}}(t_k)]\}, \tag{B5}$$

where $\mathbf{K}_k$ is the gain matrix and the assimilation has the character of an impulsive forcing term.

The perturbative equations of the *forced* system (B5) are

$$\frac{d\delta\hat{\mathbf{x}}}{dt} = \left\{\mathbf{M}[\hat{\mathbf{x}}] - \sum_k \delta(t - t_k) \cdot \mathbf{K}_k \cdot \mathbf{H}_k[\hat{\mathbf{x}}(t_k)]\right\} \cdot \delta\hat{\mathbf{x}}. \tag{B6}$$

If the observation distribution and assimilation are continuous in time, the forced system evolution becomes

$$\frac{d\hat{\mathbf{x}}}{dt} = M[\hat{\mathbf{x}}] + \mathbf{K}(t) \cdot \{\mathbf{y}^o(t) - H[\hat{\mathbf{x}}(t)]\}, \tag{B7}$$

where $\mathbf{K}(t)[\approx(1/\delta t)\mathbf{K}_k]$ represents the time distribution of the gain matrix.

The evolution of the perturbations of the continuously forced system (B7) is given by

$$\frac{d\delta\hat{\mathbf{x}}}{dt} = \{\mathbf{M}[\hat{\mathbf{x}}] - \mathbf{K}(t) \cdot \mathbf{H}[\hat{\mathbf{x}}(t)]\} \cdot \delta\hat{\mathbf{x}}. \tag{B8}$$

If the real system is chaotic, the divergence of nearby trajectories is reproduced by the perturbative equations (B3) of the free solution. In the data-forced system, the assimilation constrains the trajectory to remain close to the truth and counteracts the tendency of perturbations to amplify. As a result of the stabilizing effect of the observational forcing term in the linearized equations (B6) or (B8), the perturbations may either grow less rapidly or decay.

REFERENCES

Bennett, A. F., 2002: *Inverse Modeling of the Ocean and Atmosphere.* Cambridge University Press, 256 pp.

Bergot, T., G. Hello, A. Joly, and S. Marlardel, 1999: Adaptive observations: A feasibility study. *Mon. Wea. Rev.,* **127,** 743–765.

Berliner, L. M., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations: A feasibility study. *J. Atmos. Sci.,* **56,** 2536–2552.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.,* **56,** 1748–1765.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Dee, D., S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating covariances in distributed systems. *IEEE Trans. Autom. Control,* **30AC,** 1057–1065.

Evensen, G., and P. J. van Leewen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

Gelaro, R., R. Langland, G. D. Rohaly, and T. E. Rosmond, 1999: An assessment of the singular vector approach to targeted observing using the FASTEX dataset. *Quart. J. Roy. Meteor. Soc.*, **125,** 3299–3328.

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dyn. Atmos. Oceans,* **13,** 171–218.

——, 1997: Advances in sequential estimation for atmospheric and oceanic flows. *J. Meteor. Soc. Japan,* **75,** 289–304.

——, and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.,* **33,** 141–266.

——, and R. Todling, 1996: Tracking atmospheric instabilities with the Kalman filter. Part II: Two-layer results. *Mon. Wea. Rev.,* **124,** 2340–2352.

——, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods,* L. Bengtsson, M. Ghil, and E. Kallen, Eds., Springer Verlag, 139–224.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D-variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

Hansen, J. A., and A. L. Smith, 2000: The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.,* **57,** 2859–2871.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 796–811.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75,** 181–189.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 314 pp.

——, and Z. Toth, 1994: Removing growing errors in the analysis. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 212–215.

Lacarra, J., and O. Talagrand, 1988: Short range evolution of small perturbations in a barotropic model. *Tellus,* **40A,** 81–95.

Langland, R. H., R. Gelaro, G. D. Rohaly, and M. A. Shapiro, 1999: Targeted observations in FASTEX: Adjoint-based targeting procedures and data impact experiments in IOP 17 and IOP 18. *Quart. J. Roy. Meteor. Soc.,* **125,** 3241–3270.

Lorenz, E. N., 1984: The local structure of a chaotic attractor in four dimensions. *Physica D,* **13,** 90–104.

——, 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Vol. 1, Reading, United Kingdom, ECMWF, 1–18.

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

Montani, A., A. J. Thorpe, R. Buizza, and P. Unden, 1999: Forecast skill of the ECMWF model using targeted observations during FASTEX. *Quart. J. Roy. Meteor. Soc.,* **125,** 3219–3240.

Morss, R. E., K. A. Emanuel, and C. Snyder, 2001: Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.,* **58,** 210–232.

Navon, I. M., 1997: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Oceans,* **27,** 55–79.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.,* **55,** 633–653.

Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.,* **86,** 5878–5881.

Pu, Z. X., and E. Kalnay, 1999: Targeting observations with the quasi-linear inverse and adjoint NCEP global models: Performance during FASTEX. *Quart. J. Roy. Meteor. Soc.,* **125,** 3329–3337.

——, ——, J. Sela, and I. Szuniog, 1997: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Wea. Rev.,* **125,** 2479–2503.

Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. *Bull. Amer. Meteor. Soc.,* **77,** 953–961.

Szunyog, I., Z. Toth, K. A. Emanuel, C. Bishop, J. Woolen, T. Marchok, R. Morss, and C. Snyder, 1999: Ensemble based targeting experiments during FASTEX: The impact of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.,* **125,** 3189–3218.

Todling, R., and M. Ghil, 1994: Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.,* **122,** 183–204.

Trevisan, A., and F. Pancotti, 1998: Periodic orbits, Lyapunov vectors, and singular vectors in the Lorenz system. *J. Atmos. Sci.,* **55,** 390–398.

Uboldi, F., and M. Kamachi, 2000: Time–space weak-constraint data assimilation for nonlinear models. *Tellus,* **52A,** 412–421.