Adaptive Observations: A Feasibility Study

THIERRY BERGOT
Météo-France, Centre National de Recherches Méteorologiques, Toulouse, France

GWENAÈLLE HELLO
Météo-France, Service Central d’Exploitation de la Méteorologie, Laboratoire de Prévision, Toulouse, France

ALAIN JOLY
Météo-France, Centre National de Recherches Méteorologiques, Toulouse, France

SYLVIE MALARDEL
Météo-France, Ecole Nationale de la Méteorologie, Toulouse, France

(Manuscript received 16 October 1997, in final form 17 June 1998)

ABSTRACT

The feasibility of the recently proposed concept of adaptive observations is tested on a typical case of poorly forecast North Atlantic cyclogenesis. Only numerical tools are employed, no special observations. Although based on simulated data, this study addresses both theoretical and practical problems of adaptive observations. In the first stage of this study, the role of the data assimilation processes is neutralized; the correction is done by forcing correct continuous fields within the target area. These experiments prove that it is necessary to correct the projection of the initial errors on the first unstable plane (the first two leading singular vectors) in order to significantly improve the forecast. These results also clearly demonstrate that the quality of the initial conditions on a limited, but quite large, area could be a major factor influencing the forecast quality.

In a second stage, the focus is on operational aspects. The correction is done through the assimilation of a discrete set of simulated profiles using a 3DVAR analysis system. This leads to studying the impact of the assimilation scheme and to testing different sampling strategies. These experiments suggest that the concept of adaptive observations shows great promise in situations comparable to the one studied here. But the current assimilation systems, such as 3DVAR, require that all the structure of the target has to be well sampled to have a significant beneficial effect; sampling only the extremum does not suffice.

1. Introduction

The accurate forecast of rapidly developing cyclones is probably the ultimate dream of most forecasters and of their institutions. This is particularly apparent, for example, in the midcentury review of progress in forecasting written by Douglas (1952). In this somewhat pessimistic article, Douglas notes that all attempts to change the forecast methodology or means (use of Norwegian polar front, having ships in midocean, etc.) were proposed as the ultimate solution to this plaguing problem and failed. This was before the advent of numerical weather prediction, which, naturally, raised similar hopes. However, in spite of indisputable progress, the numerical forecast of rapid cyclogenesis remains unreliable (see Jarraud et al. 1989 for an illustration).

A key turning point toward the understanding of this vexing situation has been provided by the work of Lorenz (1963). He exemplified the extreme sensitivity to initial conditions that appears to be inherent to nonlinear dynamical systems such as those representing part or all of atmospheric dynamics. The far-reaching consequences of Lorenz’s work and of all the studies that it has triggered clearly contradict the project of forecasting the weather in general and of using a deterministic set of equations to do it in particular. Do we have to accept it and throw up our hands each time an unpredicted storm hits a coast or can we find some alternatives?

The history of science often shows examples of how a given limitation has been turned into a powerful predictive tool. A finer understanding of the limits of pre-
dictability in the specific case of cyclogenesis results from the work of Farrell (1989, 1990). He points out that the theoretical paradigm of cyclogenesis, based on normal mode instability, has serious limitations. In particular, it is not able to explain properly the timescale of full development. The timescale provided by this classical approach has been used by Eady (1949) to explicitly set up some limitations to the ability of forecasting cyclones. Farrell notes that singular vectors, rather than normal modes, appear to offer better models of cyclone development, at least with respect to timescale. Farrell’s results can be read in two ways. On the one hand, the amplification of a singular vector is much larger than that of a normal mode, and consequently the predictability is even more reduced than Eady expected. In that sense, it explains why forecast failures can still happen in the short range even with present-day systems. On the other hand, singular vectors, although responsible for the largest error growth, are calculable structures.

Such was the position with respect to understanding cyclogenesis when the early planning stages of the Fronts and Atlantic Storm Track Experiment (FASTEX) (Joly et al. 1997) field experiment on North Atlantic cyclones took place. Some of its objectives were precisely to address the predictability of cyclones [see Fig. 1 of Joly et al. (1997) for a recent example of this problem]. It is in the course of the planning of this project that the idea of “adaptive observations” emerged (Snyder 1996).

Here, the basic concept is summarized. The lack of predictability is not uniform over the globe, but strongly flow-dependent, and even local-flow dependent. The singular vectors that seem to explain rapid error growth in situations conducive to cyclogenesis are not only calculable but also relatively local. The idea is to concentrate measurements in the area where the threat to forecast quality is the largest, for example, where the singular vectors can start their amplification. Given a proper use of the data collected in this way, the uncertainty should be reduced to a minimum in critical areas. It is adaptive in the sense that the location of these measurements varies from day to day. It can also be interpreted as meaning that the present organization of observation collection, which aims at a rather uniform and widespread description, may not be the proper approach to the prediction of definite features, such as a cyclone.

The adaptive observations approach is presented in mathematical terms in section 2. Then, in section 3, the role of the data assimilation is neutralized and the best possible behavior of the various strategies is obtained by substituting “exact continuous” fields in the target area. The problem of selecting a realistic discrete sampling is addressed in section 4, prior to some concluding remarks.

2. Methodology
a. Approach

The uncertainty of forecasts made with current NWP models arises from uncertainty caused by two distinct error sources, namely, from errors in the specification of the initial conditions as well as from imperfections in the model formulation. Current understanding is that within present NWP models, the initial condition errors play a more important role in the initial stage of the forecast (i.e., 2 days) at least in the extratropical regions, whereas the model errors become increasingly more important after, beyond several days (see, e.g., Rabier et al. 1994 or Pu et al. 1997). We focus here on short-term forecast errors and restrict our attention to a perfect model situation. For this purpose, two different forecasts are computed from the same model (the French operational ARPEGE model) but start with different initial conditions (see Fig. 1). The two forecasts only differ by the initial conditions at the target time, that is, the time of the calculation of the sensitive areas. And consequently, the differences between the two forecasts at final time are only, by construction, a consequence of the initial conditions difference.

One of the two forecasts is the reference experiment (which will be referred to as REF in the following sections) and the other one is considered to be a failure case (it will be referred to as BAD in the following sections). REF and BAD are actually two different real operational forecasts interrupted at target time (REF is
FIG. 2. Difference between the REF and BAD initial conditions (target time: 1200 UTC 5 Feb 1996) for the temperature at (a) 700 hPa and (b) 300 hPa. Units: K; contouring interval, one unit. Positive values, solid line; negative values, dashed line.

a 36-h forecast at target time and BAD is a 12-h forecast). The choice of REF was determined by the fact that it is indeed the forecast nearest to reality. But for this theoretical study, the nature of REF or BAD is not crucial. The important point is that in the following sections the REF trajectory will be considered as the reality and the differences between REF and BAD are considered to be representative of analysis errors. This enables us to assess the true state of the atmosphere. Initial and 48-h forecast differences are exactly known and are interpreted as “errors.” This allows us to test the success of the adaptive observations method in correcting initial errors. The main assumption, therefore, is that the model is “perfect” and can realistically reproduce the real atmosphere. This assumption is currently used in sensitivity studies (see, e.g., Rabier et al. 1996), and recent experiments suggest that the forecast errors more often arise from errors in the initial conditions than from errors in the model formulations.

Figure 2 shows the difference between the BAD and REF initial conditions (at target time) for temperature at the 700- and 300-hPa levels. Significant differences at all levels are observed over the Atlantic Ocean. It is very difficult to subjectively define the impact of this or that part of the difference fields. One can also observe the magnitude of this error field is within the bounds of uncertainties in operational analysis systems over data-sparse areas such as oceans. As shown in Fig. 3, the two operational forecasts are very different. At verifying time, REF develops a deep low (966 hPa) located south of Ireland, while BAD develops a weak low (993 hPa) located inside the Bay of Biscay. There is, in this case, both a large amplitude and a large phase error. The aim of the study is to correct BAD with the help
Fig. 3. (a) REF and (b) BAD forecasts at final time (1200 UTC 7 Feb 1996). Mean sea level pressure—units, hPa; isocontour every five units.
of adaptive observations in order to minimize the distance between the two forecasts.

The atmospheric circulation during this period was fairly zonal across the North Atlantic. This case belongs to the category of severe storms that reach the European coast somewhat unexpectedly and that are misforecasted. They do not happen on any regular basis; they are not common. This series contains such memorable cases as the FASTNET storm of 15 August 1979, the “great storm” of 15 October 1987, the Ile de France storm of 3 February 1990, the IRIS case of 6 September 1995, or more recently the sequence of two storms that occurred in western Europe on 24 and 25 December 1997. All of these cases are responsible either for considerable loss of property (large economic impact) or unacceptable loss of life (understandably leading to a large media impact). In other words, the case considered in this study, although unique, is the one for which targeting has to be shown to be effective, prior to any multivariate analysis such as the one that can now be undertaken using the FASTEX dataset.

As a first step, the problem is considered from a theoretical standpoint. Sensitive areas are determined and the most efficient way to correct the forecast error is pursued. As a second step, we consider the problem from a more practical and operational point of view looking for the most efficient way to sample the sensitive areas with pseudo-observations. The main goal is to assess the feasibility of adaptive observations in an operational context.

b. Adjoint techniques for sensitivity analysis

In the following section, the main characteristics of the adjoint technique for the adaptive observations are reviewed. A complete set of references on the use of adjoint models in meteorology is given by Courtier et al. (1993). The primitive equation model used in this study is the ARPEGE/Integrated Forecasting System (IFS) model developed through cooperation between Météo-France and the European Centre for Medium-Range Weather Forecasts (see Courtier et al. 1991). The resolution of the tangent linear and adjoint model consists of 19 vertical levels and a T63 triangular truncation (T63L19). No diabatic effects are considered during the linear and adjoint integration [but horizontal and vertical diffusion and a surface drag are introduced, as in Buzzi (1993)]. The linear and tangent integrations are performed in the vicinity of a diabatic trajectory originating from a T63L19 run of the full model. This trajectory is truncated at T21 to keep only the large-scale features. This enables us to separate the uncertain small-scale features (seen as “perturbations”) and the large-scale features (seen as the “trajectory”). Generally, the large-scale features of the flow are better forecasted than the synoptic features and consequently the truncated trajectory allows us to reduce the trajectory sensitivity of the adjoint products. As in Rabier et al. (1992), the adjoint of the nonlinear normal mode initialization is performed at the end of the adjoint integration. The integration of the linear–adjoint model was not performed over periods longer than 48 h, as this has been shown to be a reasonable time limit for the validity of the tangent-linear hypothesis (Laccara and Talagrand 1988).

1) Expression of gradient fields

Let us denote $X$ the state vector of the model, and $J(X)$ a scalar quantity expressing a diagnostic function computed from $X$. We want to quantify which perturbations at the initial time ($\delta X_0$) are most likely to create significant changes to the diagnostic function at final time, $J(X_f)$. The change of the diagnostic function $J$ is

$$
\delta J = \left( L^* \delta J / \delta X ; \delta X_0 \right),
$$

(1)

where $L$ and $L^*$ are, respectively, the tangent-linear and adjoint operators and $\left( \cdot ; \cdot \right)$ is the canonical scalar product. Here, $(L^* \delta J / \delta X)$ is the gradient field with respect to the initial conditions and can be interpreted as a sensitivity field (the change in $J$ at final time caused by a change in the initial conditions). This method has produced results in real situations (Errico et al. 1993; Langland and Rohaly 1996) as well as in idealized situations (Rabier et al. 1992; Langland et al. 1995).

In this study, the diagnostic function used is the enstrophy inside the area of interest. Our choice is inspired by recent climatological work (Ayrault et al. 1995), as well as work on sensitivity of an idealized frontal wave (Langland et al. 1995). This choice is definitely event oriented rather than “forecast-error” oriented.

2) Singular vectors

On the other hand, the singular vectors (SV) are the most unstable perturbations (at target time) growing in a finite time interval and for given norms. So, the SVs indicate areas where initial errors, due to the lack of data, may grow superexponentially. For two given scalar products, $\left( \cdot ; \cdot \right)_{E_1}$ at initial time and $\left( \cdot ; \cdot \right)_{E_2}$ at final time, the norm of the perturbation at final time is

$$
\left\langle \delta X_f ; \delta X_f \right\rangle_{E_2} = \left\langle L^E_2 L \delta X_0 ; \delta X_0 \right\rangle_{E_1},
$$

(2)

where the operator $L^E_2$ is the adjoint of the operator $L$ following the norms $E_1$ and $E_2$ [this operator will be explained in Eq. (5)].

The square root of the eigenvalues of the matrix $L^E_2 L$ are called the singular values ($\lambda_i$) and the eigenvectors are the singular vectors.

Let us define the weighting factors of the norms at the initial time ($E_i$) and at final time ($E_f$):

$$
\left\langle X_0 ; X_0 \right\rangle_{E_1} = \left\langle X_0 ; E_1 X_0 \right\rangle
$$

$$
\left\langle X_f ; X_f \right\rangle_{E_2} = \left\langle X_f ; E_2 X_f \right\rangle.
$$

(3)

The adjoint of $L$ following the norms $E_1$ and $E_2$ is given by
$$L^*_2 = E_1^* L^* E_2,$$

(4)

where $L^*$ is the adjoint operator of $L$ for the canonical scalar product.

It is of considerable interest to be able to identify perturbations whose norm is maximized over an area of interest. This can be achieved by defining a so-called local projection operator, $P$ (Buizza 1994). The application of this projection operator to the vector $X$ sets the vector $Y$ to have zero gridpoint value outside the area of interest. The operator $L$ in Eq. (2) becomes

$$L = PL.$$  

(5)

The calculation of singular vectors and singular values uses the iterative Lanczos algorithm. The numerical computations are performed by first computing the eigenvalues ($\lambda_i$) and eigenvectors ($W_i$) of the lower-dimensional dual-product operator $LL^*_2$. The eigenvectors of this operator are related to the singular vectors $V_i$ through the relation

$$V_i = L^*_2 W_i / \lambda_i.$$  

(6)

In the present study, computations have been made with an energy-based scalar product at the initial time ($E_1$). This energy scalar product is the same as the scalar product employed to study weather predictability (Molteni et al. 1996). At the final time, the scalar product used is the enstrophy ($E_2$). At this time, the projector operator has two effects: it sets the model variables, except the vorticity, to have zero value and it sets the vorticity to have zero value outside the area of interest. The state vector at final time only contains the vorticity inside the area of interest and has a small dimension. This small dimension explains why the resolution of the dual problem is more efficient than the direct one.

The use of singular vectors is intimately associated with the notion of predictability, specifically with the estimation of the growth of the forecast error (Lorenz 1965). But some controversial issues remain concerning the choice of the metrics. In fact, the singular vectors are dependent on the choice of the inner product used in Eq. (4). If the choice of this metric is arbitrary, then so also are the singular vectors. The appropriate metric for atmospheric predictability study is the error covariance metric employed in the analysis. Unfortunately, for atmospheric predictability, this covariance error metric is not well known and is not readily obtained from operational data assimilation schemes. But recent work shows that the use of an energy-based metric, as an approximation to the true analysis error covariance metric, is appropriate for the study of atmospheric predictability (Palmer et al. 1998; Gelaro et al. 1998).

c. Application of the adjoint techniques to our case study

For this study, the time period for the adjoint calculations is 48 h. This leads to a target time of 1200 UTC 5 February 1996 and a verifying time of 1200 UTC 7 February 1996. The area of interest is the geographical area ($S$) defined by 40°–60°N, 20°–0°W. The trajectory used for the calculation is that of one of the BAD forecast. The REF trajectory is not supposed to be known beforehand; only BAD is available under operational conditions. The maximum values of the singular vectors (Figs. 4a,b) as well as the ones of the gradient fields (Fig. 4c) are found in the low atmosphere around 700 hPa. The upper-level values are smaller and vertically tilted. This classic remark points out the fact that perturbations of the initial structures in the low atmosphere are potentially more efficient than comparable perturbations in the upper atmosphere. On the horizontal, one can find three principal elongated structures that present a southwest/northeast tilting. The first one is located east of Newfoundland, the second one crosses Newfoundland, and the last one is along the east coast of the United States but located farther south. It is noticeable that the two singular vectors are on quadrature in geographical space (the zero of one corresponds to the extremum of the other) and have very similar singular values ($6.3 \times 10^{-5}$ m$^{-1}$ and $6.0 \times 10^{-5}$ m$^{-1}$). These first two leading singular vectors define the first unstable plane. The gradient fields also show the same sensitive area (same tilting and same location). It appears to be very close to the two first singular vectors. This can be expected in a highly unstable situation (Rabier et al. 1996; Horanyi and Joly 1996). One can finally notice that the area thus delimited is quite large.

To test the trajectory sensitivity of the target areas, we have done the same calculations using the REF trajectory. The results are very close to the previous ones based on the BAD trajectory. For example, 92% of the energy of the first unstable BAD plane (i.e., based on the BAD trajectory) project on the first unstable REF plane (i.e., based on the REF trajectory). It is remarkable that the first two leading singular vectors are quite similar using the BAD trajectory or the REF one. Such behavior is probably crucial to the success of the technique but in principle depends on the large-scale properties of the flow only. For the case studied, the large-scale features of the flow are well forecasted by REF or BAD.

3. Forecast correction with continuous fields and low-resolution model

a. Presentation of the experiments

In the present stage of this study, a low-resolution model is used for all the forecast runs. The nonlinear model integrations are performed with the ARPEGE/IFS model at resolution T63L19. The reference low-resolution forecast REF (Fig. 5a) started from the reference initial conditions shows a deep low in the forecasted relative vorticity field near Ireland (972 hPa). On the other hand, the low-resolution forecast (Fig. 5b)
Fig. 4. (a) SV1 and (b) SV2 fields for temperature at 700 hPa valid for 1200 UTC 5 Feb 1996. Units, K; contouring interval, 0.005 units. The gradient sensitivity field for temperature at 700 hPa (c); units, $10^{-9}$ s$^{-2}$ K$^{-1}$; contouring interval, one unit. Positive values, solid line; negative values, dashed line.
FIG. 5. Relative vorticity forecast at 850 hPa for the final time (1200 UTC 7 Feb 1996) with the low-resolution model. (a) The reference forecast REF; (b) the poor forecast BAD. Units, $10^{-5}$ s$^{-1}$; contouring interval, two units (positive values only).
started from the initial condition BAD shows an elongated low, with a minimum of 984 hPa inside the Bay of Biscay. The maximum value of the vorticity at 50°N between 5° and 10°W is $19 \times 10^{-5}$ s$^{-1}$ for REF and $8 \times 10^{-5}$ s$^{-1}$ for BAD. At low resolution, the corresponding behaviors of BAD and REF still remain the same for the cyclone location. But the intensity of cyclogenesis is stronger at high resolution. This point is discussed in more detail at the beginning of section 4.

In this section, the correction is done by imposing continuous fields. This enables us to ignore the assimilation accuracy problems or, in other words, ensures perfect observations and assimilation of the correct values. The continuous correction consists of replacing the poor initial conditions with the reference ones inside geographical masks. These masks, defined in different ways, represent the “target” to be sampled. The continuous correction produces new initial conditions, and a new low-resolution 48-h forecast is computed. Finally, the forecast improvement is assessed together with the efficiency of the initial condition correction. Two types of correction within the mask have been considered.

b. “Barotropic” correction

The first type is called the “barotropic” correction: the geographical extension of the masks is the same at all vertical levels (vertical masks). These masks are defined by using the gradient sensitivity field, $\partial J/\partial X$, and the errors field (difference between the reference and poor initial conditions), $\delta X$:

$$\delta J(x, y) = \sum_{\text{variables}} \sum_{\text{levels}} \left| \frac{\partial J}{\partial X} \right| \delta X,$$

where $\delta J(x, y)$ represents the critical location where the initial errors intersect the most sensitive structures. This does not necessarily correspond to the largest initial errors, but rather to the ones that can amplify very rapidly. The three masks used (Fig. 6) are associated with three different threshold values of $\delta J(x, y)$. These different threshold values are somewhat arbitrary and the important point is the size of the different masks.

The 48-h forecasts corresponding to the previous corrections are shown in Fig. 7. To look at the forecast improvement, we can consider the maximum value of the vorticity at 50°N between 5° and 10°W: $8 \times 10^{-5}$ s$^{-1}$ for the small mask, $12 \times 10^{-5}$ s$^{-1}$ for the medium mask, and $19 \times 10^{-5}$ s$^{-1}$ for the large mask. This clearly shows that the impact of the correction for the small mask is very weak. For the medium mask, it can be seen that the forecast improvement for the value of the vorticity at 50°N is of about 35%. For the large mask, the forecast is close to the reference, with a deep low south of Ireland. Only a small error is noticeable in the location of the low (of about 100 km). One can also notice that the geographical area defined by the large mask is close to the one defined by the first two leading singular vectors at low levels.

To demonstrate that the initial conditions inside this large barotropic mask are the only responsible for the forecast failure, a further experiment is performed. In this experiment, the REF initial conditions are kept outside of the large mask, and one puts the BAD ones inside the mask. The initial conditions are then correct everywhere except inside the geographical area defined by the mask. The 48-h forecast (Fig. 8) is very close to the BAD ones, with a weak low inside the Bay of Biscay. This control experiment clearly proves that the forecast failure is a consequence of the initial errors on a limited, but quite large, area.

c. “Baroclinic” correction

For the “baroclinic” correction, the masks are vertically tilted. These masks are defined, at level $l$, by different values of $\delta J(x, y, l)$:
The threshold values of $\delta J(x, y, l)$ used to define the mask are the same for all the levels. Here, $\delta J(x, y, l)$ represents the “efficiency” of the correction at one level, for all the model variables. The three masks used are presented in Fig. 9 at 700 and 400 hPa. One observes that the important features [large values of $\delta J(x, y, l)$] are around 700 hPa and that the values near the tropopause are weak. This leads to small masks and weak correction near the tropopause. As for the barotropic case, the different threshold values are somewhat arbitrary, the important point is the size of the masks.

Figure 10 shows the 48-h forecast associated with the different corrections. The maximum value of the vorticity at 50°N between 5° and 10°W is $12 \times 10^{-5}$ s$^{-1}$ for the small mask, $14 \times 10^{-5}$ s$^{-1}$ for the medium mask, and $19 \times 10^{-5}$ s$^{-1}$ for the large mask. So the baroclinic correction seems to be slightly more efficient than the barotropic correction (the forecast derived from using the small baroclinic mask has the same characteristics as the forecast obtained from employing the medium barotropic mask). For the medium mask, the forecast improvement for the value of the vorticity at 50°N is about 50%. For the large mask, the forecast is close to the reference, like for the barotropic case. For this baroclinic large mask, the correction is mainly localized in the medium to low troposphere. Therefore, this suggests that, for this case, the crucial analysis differences are only at low levels. The comparison between the initial error field (Fig. 2) and the adjoint products (Fig. 4) shows no clear correlation between the two fields. Near the tropopause, the significant error located south of Newfoundland is outside of the different baroclinic masks and, therefore, is not corrected. This does not seem to be responsible for the forecast failure. At low levels, the important error located south of Newfoundland is partially (but not totally) corrected by the different experiments. But the differences between the medium and large mask experiments prove that there are crucial small-amplitude errors in high sensitivity areas, farther south along 35°N.

\[ \delta J(x, y, l) = \sum_{\text{variables}} \left| \frac{\partial J}{\partial X_i} \right| \delta X_i. \]  

(d) Projection of the initial error on the unstable subspace

In order to study the dependence of the forecast improvement on the correction to the initial conditions, the projection of the initial error on the unstable subspace defined by the singular vectors is examined. This projection is defined by the absolute value of the energy scalar product between the singular vector ($SV_i$) and the initial error ($\delta X_j$):
\[ \alpha_i = \| [\delta X_j; \mathbf{SV}_i] \|, \]  

(9)

where the initial error is defined as the difference between the initial conditions of experiment \( j \), \( \text{EX}_j \), and the REF ones:

\[ \delta X_j = X_{\text{EX}_j} - X_{\text{REF}}. \]  

(10)

The results of this projection of the initial errors on the unstable subspace are presented in Table 1. The first notable result is that the best forecasts (large barotropic and baroclinic corrections) correspond to the weakest projection of the initial error on the first unstable plane, defined by the first two singular vectors \( \mathbf{SV}_1 \) and \( \mathbf{SV}_2 \). These two singular vectors have the two largest singular values and they have the same order of magnitude. One can also notice that these good forecasts still contain significant error projected on \( \mathbf{SV}_3 \). This singular vector and the following ones seem to have a weaker impact on the forecast improvement. For the medium baroclinic correction, the projection on \( \mathbf{SV}_1 \) is small, but a significant initial error on \( \mathbf{SV}_2 \) remains. This correction leads to a forecast improvement but not to a good forecast. So these experiments seem to prove that it is necessary to correct all of the initial error projecting on the first unstable plane to really improve the forecast.

To demonstrate this conclusion, the impact of the errors projecting on the singular vectors is tested. In a new experiment, called T1, a linear combination of the first two singular vectors is added to the good initial conditions REF. This creates poor initial conditions with an error strictly limited to the first unstable plane. If \( \delta T1 \) is defined as the initial error of experiment T1, and \( \delta X \) as the initial error of BAD, then

\[ \langle \delta T1; \mathbf{SV}_i \rangle_{\text{EI}} = \langle \delta X; \mathbf{SV}_i \rangle_{\text{EI}} \quad \text{for } i = 1 \text{ and } 2 \]

\[ \langle \delta T1; \mathbf{SV}_i \rangle_{\text{EI}} = 0 \quad \text{for } i > 2. \]  

(11)

The 48-h forecast beginning from this initial condition is presented in Fig. 11a. The first observation is that this very weak perturbation (see Table 2) leads to a large forecast error. The maximum value of the vorticity at 50°N between 5° and 10°W is \( 19 \times 10^{-5} \text{ s}^{-1} \) instead of \( 19 \times 10^{-5} \text{ s}^{-1} \) for REF. So this experiment demonstrates that the initial error on the first unstable plane is responsible for more than half of the forecast error, 48 h later. It is also important to know the impact of the error projecting on the following singular vectors. In experiment T2, a linear combination of the five first singular vectors is added to the good initial conditions REF so that the error is only confined to these five singular vectors:

\[ \langle \delta T2; \mathbf{SV}_i \rangle_{\text{EI}} = \langle \delta X; \mathbf{SV}_i \rangle_{\text{EI}} \quad \text{for } i = 1, \ldots, 5 \]

\[ \langle \delta T2; \mathbf{SV}_i \rangle_{\text{EI}} = 0 \quad \text{for } i > 5. \]  

(12)

The 48-h forecast for this experiment is presented in Fig. 11b. One can observe that this forecast is very close to the T1 one. The maximum value of the vorticity at
FIG. 9. The baroclinic masks used for the different experiments at levels 700 (grayscale and solid arrows) and 400 hPa (solid line and dot arrows): a small mask, b medium mask, and c large mask. [The threshold values associated with these masks are 15, 35, and 60, respectively.]

50°N between 5° and 10°W is $14 \times 10^{-5}$ s$^{-1}$, as in the previous experiment with two singular vectors. This is not very surprising given the singular values spectra. The third and following singular vectors have a very weak effect on the error growth.

In conclusion, all these experiments prove the dominating effect of the first two singular vectors and demonstrate that it is necessary to control the initial error on the first unstable plane to really improve the forecast. They also show that a wide spectrum is needed if a sufficient correction is to be obtained, which may not be practical. The following SVs have a much weaker efficiency.

4. Forecast correction with discrete data

a. Comparison with the low-resolution context

The question is now how to represent the sensitive area in an operational context. The continuous fields are replaced by observations: discrete sets of vertical profiles. This leads first to testing the representativeness and efficiency of the assimilation schemes, and second to testing sampling strategies, due to both the large size of the sensitive area and the high-frequency repartition of the sensitive signal. We only focus here on operational aspects. This is the reason why the high-resolution model used now is the operational model at Météo-France. The ARPEGE/IFS model (Courtier et al. 1991) used at Météo-France has 27 vertical levels and a T149C3.5 truncation (i.e., a T520 truncation near the pole of interest, which is located over France, and a T42 truncation at the opposite pole). Because adjoint calculations require the complete knowledge of the trajectory and are thus demanding both in terms of memory and computing time, a low-resolution model is used to locate the sensitive areas at target time (the same one as in the previous section). This is not too detrimental because we are looking at synoptic features and because the adjoint model has a minimalist physics package preventing any access to mesoscale processes. Nevertheless we have to verify that the trajectory of the model at low resolution and the one at high resolution keep the same characteristics.

At the verifying time, as shown by Fig. 12 and 3, differences between the low-resolution model and the high-resolution one exist but are acceptable for our purpose. The two forecasts differ by 6 hPa for the REF experiment and by 9 hPa for the BAD experiment. For both resolutions, however, the location of the low varies only a little: the differences between BAD and REF still remain the same, a deep low located south of Ireland for REF and a small one located inside the Bay of Biscay for BAD. One can then consider that gradient fields and singular vectors that are calculated at low resolution also depict the sensitive areas at target time for the high-resolution model.

b. Methodology

In this section, $\sigma$ denotes the area in which the initial conditions of BAD are going to be modified with the help of adaptive observations. The area $\sigma$ is defined as the part of the atmosphere indicated by the first unstable plane as was shown in the previous section. It is important to note that $\sigma$ can be computed entirely on the forecast trajectory, unlike the theoretical masks. It was also shown in section 3 that $\sigma$ is quite an important piece of the atmosphere. The operational question is then to know how to represent the $\sigma$ area as efficiently as possible within a fixed set of observations. Thus several sampling strategies employed to measure the state within $\sigma$ are tested. This is a crucial problem for the operational feasibility of adaptive observations.

To create new initial conditions, an analysis is per-
performed. The fields of BAD at target time are used as guess fields. The observations are vertical columns extracted from the REF fields at target time. The assimilation scheme used is a three-dimensional variational one, which has been operational at Météo-France since May 1997. The Météo-France three-dimensional variational scheme (Moll and Boutrier 1995) uses the incremental method. This means, in particular, that the analysis is performed at a lower resolution than that of the model. The analysis currently uses a T95 truncation, 27 vertical levels, and a stretching factor of 1. A discrete set of profiles is extracted from REF at target time to make the pseudo-observations. These profiles are then assimilated as TEMP messages, with the observation error variances operationally used for this type of message. The experiments differ from one another by the way the horizontal and vertical distributions of the discrete set of profiles are considered.

c. Experiments

Different sampling strategies are tested. Some strategies have to do with the horizontal distribution of the observations. The objective is to know if all the structure of the sensitive area is important (experiments called “detailed structure”) or if only the extremum values are needed (experiments called “extremum only”). Other strategies deal with the vertical distribution of the ob-

| Table 1. Projection [$\alpha_i$, see Eq. (9)] of the initial error on the unstable subspace defined by the five first singular vectors (SV). The singular values are in units of $10^{-2}$ m$^{-1}$. The initial error at target time is defined in Eq. (10). |
|-----------------|-----|-----|-----|-----|-----|
| SV              | 1   | 2   | 3   | 4   | 5   |
| Singlular value | 6.3 | 6.0 | 3.0 | 3.0 | 2.6 |
| Initial error projection | 21.55 | 3.39 | 9.78 | 11.39 | 18.73 |
| Initial error projection after barotropic correction (small) | 11.36 | 5.35 | 2.28 | 10.59 | 16.77 |
| Initial error projection after barotropic correction (medium) | 4.11 | 4.91 | 3.55 | 5.81 | 6.16 |
| Initial error projection after barotropic correction (large) | 1.34 | 0.82 | 0.15 | 0.92 | 3.02 |
| Initial error projection after baroclinic correction (small) | 3.06 | 4.01 | 4.60 | 10.65 | 12.94 |
| Initial error projection after baroclinic correction (medium) | 0.46 | 3.45 | 5.29 | 7.89 | 4.84 |
| Initial error projection after baroclinic correction (large) | 1.41 | 0.10 | 7.06 | 3.51 | 2.63 |
Fig. 11. Relative vorticity forecast at 850 hPa (1200 UTC 7 Feb 1996) for the experiments (a) T1 and (b) T2. Units, $10^{-2}$ s$^{-1}$; contouring interval, two units (positive values only).
E1 2547 18 095 29 590 16 124 19 365 22 851 476 1052

Table 2. Energy of the initial correction for the barotropic and baroclinic corrections, and the energy of the initial error on the first two and first five SVs. Units: J m⁻².

<table>
<thead>
<tr>
<th></th>
<th>Barotropic</th>
<th></th>
<th>Baroclinic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
<td>Medium</td>
<td>Large</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2547</td>
<td>18 095</td>
<td>29 590</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 124</td>
<td>19 365</td>
<td>22 851</td>
<td>476</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1052</td>
</tr>
</tbody>
</table>

observations. As noticed by Rabier et al. (1994) the maximum of sensitivity and thus the extremum values of the singular vectors are found at low levels. It could be worth pondering whether only this part of the structure needs to be known. One can also observe that the structures of the singular vectors have a baroclinic tilt on the vertical. The upper-level forcing at the initial stage is an important point for such a phenomenon. It is also worth wondering if the description of the upper part of the singular vector structures has an impact on the improvement of the forecast even though the corresponding amplitudes are small. The fact that the values are smaller in the upper levels might be influenced by the use of the energy norm, which puts more impact on the low levels. Table 3 and Fig. 13 show the sampling strategies for the different experiments. The resulting forecasts can be seen in Fig. 14.

- **EX1**: Thirty observations are spread to sample all the structures of the two singular vectors in the lower part of the atmosphere. A very good improvement of the BAD forecast is obtained. The vorticity field is very close to that of REF. The forecast is not quite deep enough and the center of the low is a little too far north.

- **EX2**: Thirty observations are concentrated near the extremum values of the two singular vectors in the lower part of the atmosphere. An improvement of the BAD forecast also results from this approach. However, the forecast shows that the low located south of Ireland is not as well represented as by EX1 and the vorticity field is not as well reproduced.

- **EX3**: Thirty observations are spread to sample all the structures of the two singular vectors in all parts of the troposphere. This implies that the low-level sensitive structures are less sampled than in EX1. There is still an improvement in the BAD forecast but a weak one compared to EX1. The vorticity field is far from that of REF. One can also see that the low is too elongated, not well located, and not deep enough.

- **EX4**: Thirty observations are spread to sample all the structures of the two singular vectors in the lower part of the atmosphere and 15 observations are concentrated near the extremum values of the two singular vectors in the upper part of the atmosphere. The BAD forecast is improved but is far from that of REF. The weak improvement is disappointing considering the number of observations used.

- **EX6**: The same experiment as EX1 but the assimilated part of the observations are limited to the layer between 600 and 1000 hPa. The forecast is close to that of EX1. This proves that only the observations at low levels, where the amplitude of the singular vectors is large, are important.

As shown by the better results of the experiments that are based on observations spread over the whole structure (EX1 and EX4), sampling the extremum values of the sensitive structures is not enough to accurately reproduce the initial conditions. A high density of observations is needed to describe all the structure located where the maximum of sensitivity is found, that is, the low atmosphere. We can argue here about degradation of the EX3 experiment as compared to the EX1 forecast. With such a strategy one could produce a relatively good forecast. For the case studied, the production of adequate initial conditions does not require the description of the upper part of the structure. But adding observations of the upper level is helpful to improve the accuracy of the forecast: EX4, which presents such characteristics, is the best experiment. In this experiment, the forecast failure is reduced by a bit more than 50% (see Table 4).

In order to understand the performance of the adaptive observations for the different experiments, the initial error at target time is projected on the first unstable plane. This projection is defined by

\[ \delta^2 = \sum_{i=1,2} \langle \delta X_j; SV_i \rangle_{E1}^2, \]  

where \( \delta X_j \) is the initial error of the experiment \( j \).

The results of this projection of the initial error on the first unstable plane are presented in Table 4. The first observation is that the best experiments (EX1, EX4, EX6) correspond to the weakest component of the initial error on the unstable plane. We can also notice that all these experiments are based on “detailed structure” sampling strategy. On the other hand, the poor forecasts (EX2, EX3, EX5) impart an important part of the initial error to the first unstable plane. These results, and the results of the previous section, strongly suggest that reducing the initial error projection onto the first unstable plane leads to a significant improvement in the forecast skill. To achieve this goal, with an operational as-
Fig. 12. (a) REF T63, and (b) BAD T63 forecast at verifying time (1200 UTC 7 Feb 1996). Mean sea level pressure—units, hPa; contouring interval, five units. The T149 REF and BAD are as in Fig. 2.
TABLE 3. Descriptions of the experiments. The types of the experiments are summed up in Fig. 13. The target is the geographical area defined by the first unstable plan. In the type “detailed structure,” all the horizontal structure of the first two SVs is sample. In the type “extremum only,” the observations are concentrated near the extremum values of the first two SVs.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF</td>
<td>Good forecast</td>
<td>30 observations, low atmosphere target based</td>
</tr>
<tr>
<td>BAD</td>
<td>Poor forecast</td>
<td>30 observations, low atmosphere target based</td>
</tr>
<tr>
<td>EX1</td>
<td>Detailed structure</td>
<td>30 observations, all troposphere target based</td>
</tr>
<tr>
<td>EX2</td>
<td>Extremum only</td>
<td>15 observations, upper atmosphere target based</td>
</tr>
<tr>
<td>EX3</td>
<td>Detailed structure</td>
<td>30 observations, upper-atmosphere target based</td>
</tr>
<tr>
<td>EX4</td>
<td>Detailed structure</td>
<td>assimilated part of the 30 observations of EX1 in the 600–1000hPa layer</td>
</tr>
<tr>
<td>EX5</td>
<td>Extremum only</td>
<td>15 observations upper-atmosphere target based</td>
</tr>
<tr>
<td>EX6</td>
<td>Detailed structure</td>
<td>30 low-atmosphere target based</td>
</tr>
</tbody>
</table>

Simulation system such as 3DVAR, it seems necessary to have a high density of observations inside the target area. Partial data coverage in the target area appears to be a significant limiting factor on the effectiveness of adaptive observations.

However, the EX4 forecast, which has quite a reasonable density of observations in the target area, is not a totally perfect forecast. The analysis errors at observation points (Fig. 15) show that a significant part of the initial error remains after the analysis process. The guess field errors are clearly reduced, by a factor of 5 for the temperature at low levels, for example. At target time, however, some part of the initial errors still project significantly on the first unstable plane (Table 4). This part is comparable to the experiment with a medium barotropic mask of the previous section. It was proved there that this projection on the unstable plane is responsible for the forecast failure, and this explains why the success of EX4 is not complete. This result strongly suggests that the analysis scheme is not able to build efficient continuous fields (in the sense that their projection on the first unstable plane should be reduced to zero values), even with observations without errors, as is simulated here. A part of the forecast error is only a consequence of the quality of the analysis system; this clearly shows that the feasibility of adaptive observations also strongly depends on the assimilation scheme.

5. Summary and discussion

a. Summary

The concept of adaptive observations was tested, in a real case, as part of the preparation of the FASTEX experiment. The purpose of this concept is to locate observations on that part of the flow where small analysis errors will amplify most rapidly. The study presented here is only a preliminary theoretical work on the feasibility of such an approach and does not employ any real measurements. The conclusions are based on a single cyclogenesis case. This case is representative of the most remarkable misforecasts of recent years.

The different tests are based on two different forecasts that only differ by the initial conditions. One forecast is so close to the verification that we can assume that it is the verification. The other one is a poor forecast. The main goal of this study was to correct the poor initial conditions with the help of adaptive observations, and to test the forecast improvement. The role of model errors has not been studied, as, by construction, it perfectly represents the truth.

Two adjoint techniques were used to define the target area: the sensitivity field for an enstrophy cost function.
and singular vectors (SV) characterized by an energy norm at the initial time and an enstrophy norm at the final time. For the case studied, it is noticeable that the first two SVs and the gradient fields appear to be very close: this denotes the very unstable nature of the case. The maximum values of sensitivity are found, classically, in the low atmosphere around 700 hPa. The first two SVs are on quadrature in geographical space and have very similar singular values. Such a property defines the so-called first unstable plane. The following SVs have much weaker singular values resulting in flat spectra.

In the first stage, the problems related to the accuracy of the assimilation techniques are not taken into account; the correction is done by imposing so-called continuous fields. These continuous fields corrections consist of replacing the poor initial conditions by the reference ones inside different geographical masks. The main conclusion is that it is necessary to correct all the initial errors that project on the first unstable plane in order to significantly improve the forecast. For the case studied, the initial errors on the first unstable plane are responsible for more than half of the forecast error. The first unstable plane has a dominating effect on the forecast failure, and the following SVs have a weaker efficiency. The different experiments have also demonstrated that the quality of the initial conditions on a limited, but quite large, area is a major factor influencing the forecast quality. For the case studied, the “target” area is localized in the low atmospheric levels (around 700 hPa), but is of quite large areal extent. The area defined by the geographical location of the first unstable plane could therefore give a serious indication of the target area in real time.

Another important question related to the feasibility of adaptive observations is to know how to sample as efficiently as possible the target area with a fixed set of observations later handled through an assimilation system. This was the main goal of the second stage of this work. This is a crucial problem for the operational feasibility of adaptive observations. The correction is obtained from a discrete set of simulated profiles assimilated by a 3DVAR analysis system. In this way, different sampling strategies have been tested and the efficiency of the assimilation scheme has been assessed. This study suggests that the concept of adaptive obser-
vations shows great promise as a practical means of improving numerical weather forecasts in situations comparable to the one studied here. But the current assimilation system requires that all the structure of target fields (such as SVs) must be well sampled in order to have a beneficial effect; sampling only extrema does not suffice. This result implies that a large number of observations inside the target area is needed with the current assimilation system. But given a specific distribution of observations, the forecast could be well improved.

b. Discussion

The impact of the adaptive observations on forecasts can be explained in terms of projections of the analysis error onto the first unstable plane. One could argue that this result strongly depends on the singular values spectra. But the existence of a dominating first unstable plane seems to be a property of strong cyclogenesis. This property has also often been observed during FASTEX, for example, during the Intensive Observing Period (IOP)17 (see Fig. 16). This unstable plane could be used as a criterion to test the quality of the initial values. These first results and the first results using FASTEX data (Emanuel and Langland 1998) confirm that the impact of adaptive observations depends on minimizing the projection of analysis error onto the first leading SVs. In a more general sense, these results suggest that the assumptions and approximations made in applying the singular vector calculations (linearization, dry physics during the adjoint integration, energy norm at initial

<table>
<thead>
<tr>
<th></th>
<th>EX1</th>
<th>EX2</th>
<th>EX3</th>
<th>EX4</th>
<th>EX5</th>
<th>EX6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 - \frac{|\delta X(\text{EX})|<em>{\text{min}}}{|\delta X(BAD)|</em>{\text{min}}}$</td>
<td>53.6%</td>
<td>30.8%</td>
<td>34.1%</td>
<td>58.3%</td>
<td>20.5%</td>
<td>51.6%</td>
</tr>
<tr>
<td>$\delta^2$</td>
<td>31</td>
<td>75</td>
<td>78</td>
<td>26</td>
<td>70</td>
<td>25</td>
</tr>
</tbody>
</table>
time, enstrophy norm at final time, perfect model, etc.) are appropriate for improving the predictability of cyclogenesis. The fact that the choice of norms allows for a well-defined unstable plane is a valuable a posteriori justification of this choice.

It is also noticeable that recently proposed "nonadjoint" methods also exist for adaptive observations purposes and were used during FASTEX: for example, the Ensemble Transform Technique (Bishop and Toth 1996) and the Quasi Inverse Linear Method (Pu et al. 1997). These methods are associated with the bred (Lyapunov) vectors used for ensemble forecasting at the National Centers for Environmental Prediction (Toth and Kalnay 1997). In the breeding method, a random perturbation is repeatedly evolved (by nonlinear model) and rescaled to a specified amplitude over a relatively short cycling time. The quasi-inverse method finds a close approximation to the exact initial error that corrects the forecast error (Pu et al. 1997). The adjoint methods can be considered optimal in that they find the smallest perturbation that results in a maximum decrease of the forecast error (in the sense of the norms used). Both methods were used during FASTEX for adaptive observation purposes and a complete evaluation, and statistically meaningful comparison of the various targeting methods is in progress. The results of this study clearly suggest that the predictability problem is constrained both by the observing network and the process of assimilating and analyzing the observations to produce initial conditions. In principle, adaptive observations could revolutionize the methodology for determining initial conditions for weather prediction by interactively coupling the process of data assimilation and forecast with that of measurement. This study also suggests that the sampling strategy used for adaptive observations is also an important problem to solve. While the SVs point out the large area where adaptive observations are needed, the real issue is now to find the optimal locations where

the smallest amount of additional adaptive observation will best minimize the forecast errors.

The results of this study also suggest that a current operational assimilation system, such as 3DVAR, is not able to build efficient fields, with no initial errors on the first unstable plane relevant to the subsequent forecast period. Consequently, a nonnegligible part of the forecast error is simply due to the limitations of the analysis system. This result clearly shows that the success of adaptive observations also depends on the assimilation scheme. The deficiency of a 3DVAR system could be explained by the fact that the structure functions that distribute information are not flow dependent. We have checked that the even simpler structure functions used in optimal interpolation are less able to reduce the projection on the critical unstable plane. Recent work within the framework of an idealized atmospheric system but using a sophisticated assimilation technique with well-evolved correlation functions (Fischer et al. 1998) strongly suggests that 4DVAR might be very useful for maximizing the impact of adaptive observations. The next significant step is clearly to experiment on the series of FASTEX cases with a 4DVAR analysis scheme.

Acknowledgments. We acknowledge the Météo-France assimilation team, and particularly Jean-Noël Thépaut and Philippe Caille, for their friendly support during the 3DVAR experiments. And we want to thank Philippe Arbogast, Gérald Desroziers, and Béatrice Pouponneau for helpful discussions and computing assistance.

REFERENCES


Bishop, C., and Z. Toth, 1996: Using ensembles to identify observations likely to improve forecasts. Proc. 11th Conf. on Nu-


