Q. J. R. Meteorol. Soc. (1999), **125**, pp. 3189–3217

Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet

By I. SZUNYOGH[1]*, Z. TOTH[2], K. A. EMANUEL[3], C. H. BISHOP[4], C. SNYDER[5], R. E. MORSS[3], J. WOOLEN[2] and T. MARCHOK[2]

[1]*NCAR/MIT visiting scientist, National Centers for Environmental Prediction, USA*
[2]*General Sciences Corporation, National Centres for Environmental Prediction, USA*
[3]*Massachusetts Institute of Technology, USA*
[4]*The Pennsylvania State University, USA*
[5]*National Center for Atmospheric Research, USA*

(Received 18 May 1998, revised 8 February 1999)

SUMMARY

In this study we evaluate the performance of the Ensemble Transform (ET) technique, which is one of several targeting methods used in real time during the Fronts and Atlantic Storm-Track EXperiment (FASTEX). 'Targeted' observations were taken adaptively in those upstream areas identified in real time as most relevant for improving the initial conditions for forecasts of synoptic-scale storms developing downstream. The upstream areas were identified as regions where the effect of extra observations at a future analysis time could produce the largest decrease in the largest likely forecast error at a preselected later verification time at a given downstream location. The ET technique selects these observational areas out of a large number of possible deployment locations of observational resources via a linear transformation of an ensemble of forecasts.

The analysis and forecast effects of special targeted observations associated with seven Intensive Observing Periods (IOPs) during FASTEX were investigated. The most important result of the present study is that the ET technique, based on the National Centers for Environmental Prediction (NCEP) operational global ensemble, was able to identify upstream areas that had significant contribution to the quality of selected future downstream forecast features. Moreover, the technique could reliably distinguish between the areas of greatest contribution.

Though the overall impact of the targeted data on forecast quality is positive, there were cases when the extra data degraded the forecasts. Our analysis indicates that large amplification of ensemble perturbations from the targeted area into the verification area is a good indicator of potential forecast improvement.

KEYWORDS: Ensemble Transform (ET) technique Fronts and Atlantic Storm-Track EXperiment (FASTEX) Targeted observations

1. INTRODUCTION

Serious short-term forecast failures are localized both in time and space. Concerning the mesoscale and larger scales, these failures are

often associated with intense atmospheric development and are thought to be due more to errors in the initial conditions than to model deficiencies (Lorenz 1990). In data-sparse regions, analysis errors are not only larger than average but also dynamically better organized, since amplifying errors in the first-guess forecasts cannot be efficiently reduced (Iyengar*et al.* 1996). Experiments with an operational analysis–forecast system assimilating dropwindsonde observations taken in the north-east Pacific (Lord 1996) and with a low-dimensional dynamical system (Lorenz and Emanuel 1998) suggested that extra data in the oceanic regions in general have a positive influence on the quality of the analysis and the ensuing forecast downstream over land. Emanuel *et al.* (1996) and Lorenz and Emanuel (1998) also concluded that for their low-order model there exist operationally attainable adaptive observation strategies that are superior to the traditional procedure of taking observations at fixed locations and times only or as opportunity arises. They found that ensemble-based strategies which sample regions with the largest expected error in the first-guess forecasts perform better than those that target areas to which a pre-specified verification region is most sensitive at a given

* Corresponding author: National Centers for Environmental Prediction, EMC, 5200 Auth Road, WWB, Room 207, Camp Springs, MD 20746, USA.

forecast lead time.* Experiments with a more complex quasi-geostrophic model (Morss 1999) suggested that in certain circumstances adaptive observations have the potential to improve analyses and forecasts more than traditional observations, but that at the same time there is a non-negligible risk that in certain cases a forecast can be degraded with targeted observations.

Nonetheless, a skilful *operational targeting procedure*† that was able to reduce the uncertainty in the prediction of dramatic weather events, such as a land-falling storm or hurricanes threatening highly populated areas, would be invaluable from a socio-economic perspective. In the past few years, different techniques have been suggested for identifying the optimal regions for targeted observations. Palmer *et al.* (1998), Bergot *et al.* (1996, 1999), Langland and Rohaly (1996) and Kalnay and Pu (1997) suggested the use of sensitivity gradients and/or dominant singular vectors. Pu *et al.* (1998a) also suggested using the quasi-inverse technique developed by Pu *et al.* (1997) for targeting. All of these methods use the tangent–linear and/or the linear–adjoint version of the nonlinear forecast model.

The Ensemble Transform (ET) technique (Bishop and Toth 1996, 1998), which will be further discussed in section 3 is another objective targeting method, which is based on an ensemble of forecasts instead of tangent–linear model integrations.

Targeting for the short range can also be performed subjectively, tracking, for example, potential-vorticity anomaly patterns in a series of maps (A. J. Thorpe, personal communication). The general mathematical formulation of the objective targeting problem and some relationships between the different proposed techniques are discussed in Berliner *et al.* (1999), though there is no study available concerning the comparison between the performance of the different targeting strategies used in practice. While preliminary studies indicate that objective targeting techniques can potentially be useful for targeting observations (Pu *et al.* 1998a, b; Bergot *et al.* 1998; Bishop and Toth 1998; Gelaro *et al.* 1999), the Fronts and Atlantic Storm-Track EXperiment (FASTEX), (Joly *et al.* 1997, 1999) in January and February 1997, offered, for the first time, an opportunity to test a number of these techniques in the field. In this study we focus on those February FASTEX cases for which the ET technique was used as the primary tool for identifying supplementary observation sites.

Section 2 is a short summary of the main objectives of and the terminology used during FASTEX. In section 3 we describe the particular form of the ET method that was used during FASTEX. Section 4 contains a general description of analysis–forecast experiments, while section 5 evaluates the performance of the ET technique during FASTEX. Section 6 is devoted to the forecast-verification results, while section 7 offers some conclusions.

2. The Fronts and Atlantic Storm-Track EXperiment (FASTEX)

A detailed description of the main objectives, operation plans, observing instruments, participating institutions and scientists for FASTEX can be found in Joly *et al.*

* In their paper, a downstream geographical region is said to be sensitive if the gradient of a response function measuring the amplitude of the forecast uncertainty at a later verification time in the downstream region has non-zero magnitude at that location.

† The notion of targeting as used here should not be confused with the existing definitions of dynamical-systems theory. In dynamical-systems theory targeting means a series of perturbations applied to a nonlinearly evolving phase point (just as with observed data through the analysis cycles) in order to force it to visit the neighbourhood of a predefined point on the attractor (just as with the real state of the atmosphere at the verification time within the verification region) as quickly as possible (e.g. Bollt and Kostelich 1998). Despite the similarities the two definitions are obviously different.

(1997), while an overview of preliminary results is given by Joly *et al.* (1999). Here, we review only those aspects that are crucial to understanding the analysis–forecast experiments presented below.

The main focus of FASTEX was to study the development of frontal cyclones over the Atlantic. Two observational components were included in FASTEX: a multi-scale component designed to document the structure of mature frontal cyclones off the coast of Ireland, and an upstream component designed to document incipient cyclones and to provide targeted observations to improve downstream forecasts. The multi-scale component utilized three aircraft based at Shannon (Ireland), two of them with Doppler-radar (WP-3D of the National Oceanic and Atmospheric Administration (NOAA); Electra of the National Center for Atmospheric Research (NCAR)) and a third one with drop-sounding capabilities (UK C-130). Observations from these platforms were taken during 19 IOPs, at times when mature systems reached the Multi-scale Sampling Area (MSA) defined by the 2.5-hour ferry distance ($\approx$900 km) of the WP-3D aircraft from its base of operations at Shannon.

The upstream observing platforms included a Lear jet commissioned by NCAR, two US Air Force C-130s, and occasionally the NOAA Gulfstream-IV. The Lear jet was based at St. John's, Newfoundland (47.64°N, 52.75°W) and the flight planning for that aeroplane was coordinated by three of the authors (K. A. Emanuel, C. Snyder and Z. Toth). The flight planning consisted of two steps: (1) Identification of a forecast event at a given location and time, associated with a frontal wave or cyclone, and (2) identification of the corresponding upstream sensitive area. In order to accomplish step (1), control forecast guidance from several numerical weather prediction (NWP) centres was used. Event selection was based both on the suitability of the forecast event for sampling in the MSA and on estimates of forecast uncertainty, with the former criterion receiving greatest emphasis. The estimate of forecast uncertainty was based upon the spread in the National Centers for Environmental Prediction (NCEP) operational ensemble (Toth and Kalnay 1997). The ensemble often pointed to the high-pressure area behind the cyclones as the most uncertain area, suggesting good model performance for the cyclone itself. The generally good performance of the NCEP control forecast for the FASTEX cyclone cases subsequently justified the use of the ensemble spread as an indicator of forecast uncertainty. For step (2), targeting products from the Naval Research Laboratory (NRL), Monterey, USA; Meteo France, Toulouse, France; the European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK; NCEP, Washington, DC; and Pennsylvania State University, University Park, USA were considered. Verification of the targeted forecasts was only a secondary goal of the multi-scale observations. Because of this, only a limited number of drop-soundings are available in the verification region. To remedy the problems we will frequently extend the verification region to include rawinsonde soundings over the surrounding area.

FASTEX took place in January and February 1997. Unfortunately, during much of January a blocked-flow regime dominated the circulation over the Euro–Atlantic region. Thus, the three Lear-jet flights in January could only target the precursors of areas of large forecast uncertainty not related to strong cyclone development. Here we present analyses from seven February cases. These are identified by their Intensive Observation Period (IOP) number; IOP9, IOP11, IOP13, IOP15, IOP16, IOP17, and IOP18, out of a total of 11 FASTEX cases from February. In a subsequent study, Bergot (1999) evaluated the effects on forecasts of targeted dropsonde data released from two of the four upstream aircrafts (Lear and G-IV), using the ARPEGE analysis–forecast system of Meteo France.

3. The Ensemble Transform technique

(*a*) *General algorithm*

The ET technique (Bishop and Toth 1999) provides a rapid algorithm to estimate how forecast errors in a pre-selected verification region may change depending on the locations where extra observations are taken at an earlier time. During FASTEX, the threat of a major forecast failure was estimated by the largest likely forecast error, G, associated with an assumed analysis-error pattern that has fixed magnitude over the North Atlantic. We predict changes in G due to changes in the geographical distribution of the analysis uncertainty, or, more formally, to changes in the analysis uncertainty matrix. This algorithm computes $G(\mathbf{A}_i)$ for a series of estimated analysis-uncertainty matrices, $\mathbf{A}_i$ $(i = 1, \ldots, N)$, associated with the standard observational network augmented with hypothetical extra observations from N different locations. The optimal deployment of the observational resources is the one, $j \in [1, N]$, that leads to the largest reduction in G ($G(\mathbf{A}_j) \leqslant G(\mathbf{A}_i)$ $i = 1, \ldots, N$), or in other words, the one that most efficiently reduces the threat of a large forecast failure within the verification region.

The formal maximization problem to be solved for each possible adaptation of the observational network is as follows. Given are a verification region and a K-member ensemble of forecast perturbations, $\mathbf{x}_k(t)$ $(k = 1, \ldots, K)$, initiated at t_0, valid at a future analysis (targeting) time $t_a (> t_0)$ and at a verification time $t_v (> t_a)$. These ensemble perturbations, $\mathbf{x}_k(t)$ $(k = 1, \ldots, K)$, can either be defined as the difference between certain ensemble members or as the difference between the ensemble members and the ensemble average. For each realization, $i \in [1, N]$, of the observational network the ET technique finds the linear combination, $\mathbf{y}(t) = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_K\mathbf{x}_K(t)$, of the ensemble perturbations that maximizes the perturbation amplitude $\|\mathbf{L}\mathbf{y}(t_v)\|^2 = \langle \mathbf{L}\mathbf{y}(t_v), \mathbf{L}\mathbf{y}(t_v)\rangle$ in the verification region subject to the constraint that the perturbation has unit magnitude at the future analysis time t_a in the following sense $\langle \mathbf{y}(t_a), \mathbf{A}_i^{-1}\mathbf{y}(t_a)\rangle = 1$. $\langle ., . \rangle$ and $\|.\|$ are the Euclidean scalar product and the associated norm for grid points that span the North Atlantic including the verification region. L is the localization operator that sets $\mathbf{y}(t_v)$ to zero in regions outside of the verification area; c_k, $k = 1, \ldots, K$ are constant scalars. By using the same set of coefficients at analysis and verification times, we assume that the time evolution of the nonlinear localized ensemble perturbations, $\mathbf{L}\mathbf{y}(t_v)$, can be approximated by a linear model. Although the matrix of this model is never computed explicitly in the ET technique, its largest singular value gives the solution $G = \max \|\mathbf{L}\mathbf{y}(t_v)\|^2$ of the maximization problem. For a given set of ensemble perturbations G solely depends on $\mathbf{A}_i$, which will be different for the different adaptations of the observational network. Bishop and Toth (1999) demonstrated that if K is equal to the number of degrees of freedom in the model and a reliable estimate of the full analysis-error covariance matrix, $\mathbf{V}_i$, is available for each adaptation of the observational network the statistically optimal choice for the constraint matrix is $\mathbf{A}_i = \mathbf{V}_i$. In that case G is the largest singular value of the forecast-error covariance matrix.

However, in practice K is extremely small compared to the degrees of freedom in an operational NWP model and only static (time-independent) estimates are available for the diagonal entries of $\mathbf{V}_i$. In a typical application of the ET technique $\mathbf{A}_i$ is based on a static estimate of the diagonal analysis-uncertainty matrix, which imposes a much weaker constraint on the maximization than a full $\mathbf{V}_i$ in the idealized case above. In fact, a new constraint arises when using the ET technique with a small, finite-size ensemble

since the search for $\mathbf{y}(t)$ is constrained at both t_a and t_v to the low-dimensional subspaces spanned by the ensemble perturbations.

The computation of the constraint matrix consists of two steps. First, an estimate of the analysis-uncertainty matrix, $\mathbf{A}_0$, associated with the standard (or non-adaptive) observational network is determined. Then the impact of the hypothetical extra data is introduced into the estimate $\mathbf{A}_i$, through reducing the entries in $\mathbf{A}_0$ that were associated with locations within a particular targeted region. This rescaling strategy assumes that extra data can reduce the magnitude of the analysis uncertainty in the targeted region. In this region the maximization constraint will allow only for reduced-amplitude initial perturbations, therefore reducing G if there is a correlation between the ensemble perturbations in that region at the future analysis time t_a and within the verification region at verification time t_v.

(b) *Application during FASTEX*

The ensemble perturbation vectors were defined as the difference between the positive and the negative members of paired nonlinear ensemble forecasts. Thus, the 10-member 00 UTC operational NCEP ensemble forecast gave five independent perturbation vectors, while the four members of the 12 UTC ensemble gave two additional vectors. Regular targeting products were provided in a quasi-operational mode twice a day, using the latest-seven perturbation vectors after the operational ensemble runs had finished. However, due to computational constraints on the NCEP Cray computers, interactive calculations specifically prepared for flight-planning purposes used only the latest five NCEP perturbation vectors. The verification region for the standard calculations was a 1000 km radius disc with a standard central location at (50°N, 10°W), while in the special sensitivity calculations for flight planning the position of the disc was adjusted toward the actual location of the FASTEX cyclone as it was forecast at the time of targeting (t_a), also considering the area of largest ensemble spread for the surface pressure within the MSA. The verification variables, as well as the 'observed' variables at t_a were stream-function grid-point values at pressure levels 850, 500 and 250 hPa.

In order to assess the dependence of the ET algorithm on assumed standard analysis uncertainty, two different choices for $\mathbf{A}_0$ were tested during FASTEX. One of them was the identity matrix assuming that the ensemble is representative of the distribution of analysis uncertainty over the globe before the deployment of the extra observations. The other, presumably more realistic, choice for $\mathbf{A}_0$ was estimated by a long-term average of the root mean square (r.m.s.) difference in the rotational wind-speed between analyses from two independent analysis cycles (Iyengar *et al.* 1996). One of these analyses is the operational T126-horizontal-resolution analysis truncated to resolution T62, while the other is the NCEP T62 re-analysis. The average r.m.s. difference field was smoothed by a spectral Gaussian filter (J. Purser, personal communication) that removed the variability associated with wave numbers higher than 7–8 from the variance field (Iyengar *et al.* 1996).

We assumed that the N possible adaptations of the standard observational network were aircraft-based observations over the North Atlantic that would reduce the analysis-error variance over a disc of radius 350 km centred at a location (ϕ, λ). We imitated the impact of extra data by reducing the entries in $\mathbf{A}_0$ that were associated with grid points within the disc. In what follows we will distinguish among the estimated analysis-error matrices associated with the different adaptations of the observational network by indicating the central position of the hypothetical targeted region as $\mathbf{A}(\phi, \lambda)$ instead of using subscript i.

In cases when the analysis-uncertainty-based estimate of $\mathbf{A}_0$ (Iyengar *et al.* 1996) was used $\mathbf{A}(\phi, \lambda)$ was computed by replacing the diagonal entries associated with the disc centred at location (ϕ, λ) by the smallest diagonal element of the full matrix. In other words, we assume that the hypothetical targeted observations can reduce the analysis uncertainty to its lowest estimated value on the globe over a disc of radius 350 km.

When the unity matrix was used as $\mathbf{A}_0$ the diagonal matrix entries related to grid points within the observational region were reduced to 0.33. As expected, the use of the unity matrix generally resulted in larger sensitivity over land and close to the eastern shore than the use of the analysis-uncertainty-based estimate of $\mathbf{A}_0$. However, in most cases the two products identified the same oceanic regions as the minimum of $G(\phi, \lambda)$.

To find the position (ϕ, λ) of a target disc that would minimize G, we estimated G for a set of λ and ϕ values that spanned the North Atlantic. $G(\mathbf{A}(\phi, \lambda))$ was calculated at ten-degree resolution in the regular daily products and at five-degree resolution in the specific calculations preceding the flight decisions. This ten- or five-degree spacing between the discs over which extra observations could be taken should not be confused with the resolution of the dynamically evolving ensemble perturbations, which have a horizontal resolution of T62.

The final sensitivity products are geographical plots of $G(\phi, \lambda)$. Examples of such estimated forecast-sensitivity maps are given in Fig. 1, where $G(\phi, \lambda)$ is normalized by its maximum over the geographical domain that is shown. The forecast within the verification region is said to be *more sensitive* to the analysis at location (ϕ_1, λ_1) than at (ϕ_2, λ_2) if $G(\phi_1, \lambda_1) < G(\phi_2, \lambda_2)$.

Since the main goal of this study is to test the feasibility of the proposed real-time decision making mechanism for targeting, all results presented here are based on targeting products that were calculated and used in real time during FASTEX.

(*c*) *Potential limitations of the Ensemble Transform technique*

If all the assumptions of the ET technique were completely satisfied and it was possible to accurately predict the effect of extra data on the analysis error this method would quantify the reduction in the largest likely forecast error brought about by the extra observations taken at the different locations. However, in practice there are a few assumptions on which the method and its application are based that could potentially limit the usefulness of the technique. Firstly, it is assumed that linear combinations of ensemble forecast perturbations can be used to represent future analysis errors. We must recognize that the definitions of $\mathbf{A}(\phi, \lambda)$ for the two $\mathbf{A}_0$ used during FASTEX were reasonable but to a large extent, ad hoc choices. In general, we expect the ET technique to provide more reliable results regarding the relative distribution of predicted sensitivity ($G(\phi, \lambda)$) than the actual estimated values of forecast-error reduction. The fact that in most cases the analysis-uncertainty-based estimate of $\mathbf{A}_0$ and the identity matrix gave virtually identical results suggests that the statistical structure of the ensemble, and not the $\mathbf{A}_0$ constraint, played the dominant role in this particular FASTEX application of the method. Secondly, it may not always be valid to assume that the time evolution of the dominant error pattern from t_a to the verification region at t_v can be well approximated by linear combinations of the nonlinear ensemble perturbations. This may happen either when the ensemble perturbations are large at t_a or when the time evolution of the error pattern is associated with initially small but nonlinear instabilities.* Thirdly, it

* A nice example for an exclusively nonlinear instability process is the echo effect discussed by Vanneste *et al.* (1998) for a two-dimensional barotropic (Couette) flow.

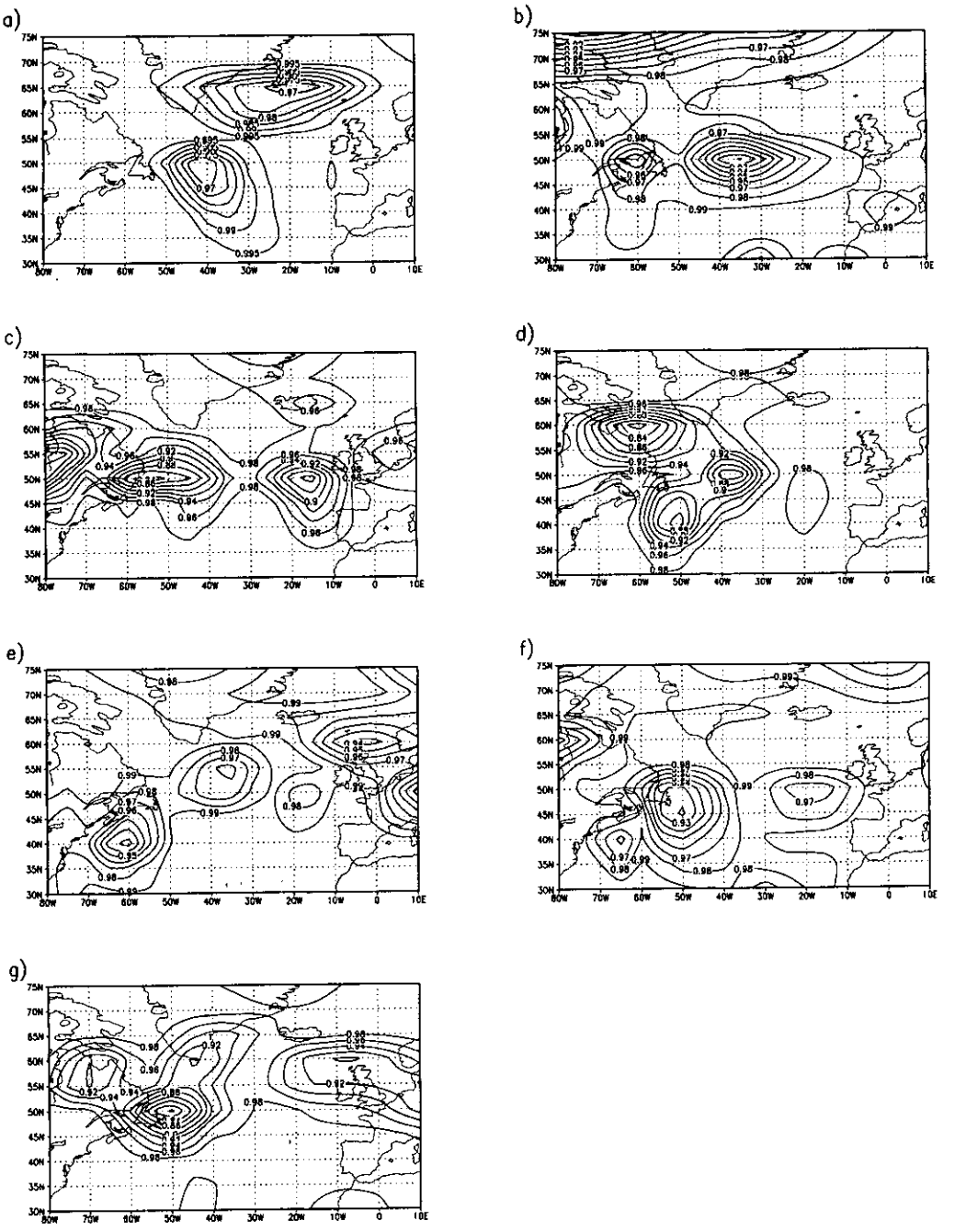

Figure 1. The sensitivity maps that were used for flight decisions in real time. The investigated cases are IOP9 (a), IOP11 (b), IOP13 (c), IOP15 (d), IOP16 (e), IOP17 (f), and IOP18 (g) (see text). The regions of the largest possible impact from extra observations are where the mapped values of $G(\mathbf{A}(\phi, \lambda))$ are minimal. See text for further explanation.

is assumed that a small set of ensemble forecasts can be used. A small ensemble may be adequate to capture the relevant dynamics of forecast-error evolution in cases of fast error growth associated with a few rapidly expanding phase-space directions. But when error dynamics is governed by a larger number of phase-space directions, characterized by similar expansion rates, the method may be seriously limited even if the linear approach is valid. In these cases, as in any other statistically based Singular Value Decomposition (SVD), the ET technique may find spurious minima of G due to the small sample size (Cherry 1996). In order to eliminate the spurious local minima of $\mathbf{A}$ related to low sample size, time series of sensitivity calculations were created for time windows 12–84, 24–84, . . . , and 72–84 hour lead times. On several occasions special products were also prepared based on 25 perturbation vectors from the operational, T159, 50-member, 12 UTC ECMWF ensemble in a way analogous to that used with the NCEP ensemble (Toth and Szunyogh 1997). The results of these runs (not shown in this paper) typically agreed closely with those made with the NCEP ensemble. Fourthly, the NCEP ensemble formulation is based on a perfect-model hypothesis, which has limitations especially when the ET calculations are done between time levels of long forecast lead times.

(*d*) *The relationship between the Ensemble Transform and the adjoint-based techniques*

The ET technique computes the dominant singular values (associated with the dominant singular vectors) of an estimated linear error-propagation matrix for a large number of different analysis-uncertainty distributions associated with different hypothetical observational configurations, to explicitly choose the configuration that most efficiently reduces the threat of a major forecast failure. Such computations, in principal, should be possible using the adjoint techniques (Barkmeijer 1998). Computational limitations, however, are prohibitive with realistic model configurations. Therefore, traditional adjoint-based sensitivity calculations determine perturbation (or 'error') patterns (singular vectors or sensitivity patterns) that can amplify the most, given fixed assumptions about analysis-error variance/covariance.

The practical choice is basically between performing only one (or a few at maximum) adjoint-based estimations associated with one (or a few) estimated analysis-error structure(s), or performing many of these calculations, but for a limited number of analysis and forecast variables and in the limited sub-space of ensemble configurations.

4. DESCRIPTION OF THE ANALYSIS–FORECAST EXPERIMENTS

(*a*) *The analysis–forecast system*

All analysis–forecast experiments presented here were carried out with the NCEP forecast system that was operational during FASTEX. This includes a T62, 28-level version of the operational NCEP Medium-Range Forecast model (MRF), (Kanamitsu *et al.* 1991) and the Spectral Statistical Interpolation (SSI) scheme (Parrish *et al.* 1997), a three-dimensionsl (3D), variational data-assimilation scheme. The T62 resolution version was chosen because a considerably larger number of experiments could be completed than at the higher, T126 resolution. As mentioned earlier, the T62 model forecasts for FASTEX cyclones verified rather well, sometimes even better than the T126 predictions. Because of the relatively short duration of the Lear-jet flights, it was possible to assimilate all targeted dropsonde data in single 6-hour analysis time windows in each of the cases studied here, without the need for cycling the analysis.

A detailed description of the SSI variational assimilation scheme can be found in Derber *et al.* (1991); Parrish and Derber (1992) and Parrish *et al.* (1997). Here we review only those features of the scheme that may have a significant effect on our targeting experiments. The analysis is a weighted average of the observed data and the short-range model forecast (first guess, background) valid at the time of observations. The interpolation weights for the observed data are inversely proportional to the square of the estimated observation errors, which are different for different observation types but fixed in time. Except in the case of the satellite temperature soundings, the observation errors are assumed to be uncorrelated. The information from the observational increment (the difference between the observation and the first guess) is spread out using the spatial structure of the background-error covariances, and thus the structure of the analysis increment (the difference between the analysis and first guess) is primarily determined by the background-error statistics. These time-independent statistics are assumed to be homogeneous and isotropic in the horizontal grid space. The covariances between the rotational component of the wind and the balanced part of the mass fields are determined by an assumed, nonlinear balance relationship (the spatial covariances between the other variables are not constrained). This means that neither the background-error statistics nor the weights for the observations in the analysis depend on the synoptic situation or the geographical location.

Since several FASTEX cases appeared to have been associated with large background errors (estimated from large ensemble spread), these approximations in the analysis scheme can lead to substantial errors in the analysis. In particular the information from the dropsonde observations is extrapolated deep into the relatively data-sparse regions via the linear balance equation without the control of additional observed data.

Another limitation of the then operational SSI scheme was that for the calculation of the observational increments, the background field was interpolated to the actual observation time only for the first three hours of the 6-hour analysis period. In the second half of the analysis period the first guess was interpolated only in space, which may introduce a phase error into the analysis of rapidly developing atmospheric systems.* Note that the above limitations of the SSI scheme are typical of current operational analysis schemes.

(*b*) *Evaluation methods*

All dropsonde data were quality controlled: observations that represented obviously unrealistic states of the atmosphere or contradicted the hydrostatic-balance condition were rejected. Other data were rejected following a subjective inspection. Neither the optimal-interpolation based (Woolen 1991), nor the complex quality control (Gandin 1990) components of the operational quality control procedure were applied to the dropsonde data because those components involve assumptions about the statistical structure of the meteorological fields (e.g. homogeneous and isotropic, spatial-correlation structures that are fixed in time) that may not be valid for fast-developing synoptic-scale storms.

Two analyses were created for each FASTEX case, one with and another without the dropsonde data from the Lear jet. All other FASTEX data were included in both forecasts, because the upstream aeroplanes were the only observing platforms used during FASTEX for which the observation times and locations were selected exclusively on the basis of objective targeting products. Taking these analysis fields, two forecasts were run until the verification time of the targeting product.

* Since the summer of 1998, the operational version of the SSI uses a 9-hour first-guess forecast which facilitates the time-interpolation for the second half of the 6-hour assimilation window, too.

For successful targeting one wishes to find a difference (which we will refer to as a 'signal') of significant amplitude between the two forecasts that reaches the verification region at the verification time. Even in this case, however, it is possible that observations at other locations could produce a signal that would reach the verification area at verifying time with an even larger amplitude. Therefore, for all IOPs we run 'mock' targeting experiments in which all data from areas adjacent to the objectively targeted regions (except for the satellite soundings) were withheld from control analyses. Two mock targeting regions were defined by the boxes $45°\text{N} \leqslant \phi \leqslant 60°\text{N}$, $57.5°\text{W} \leqslant \lambda \leqslant 70°\text{W}$ (M1 region) and $30°\text{N} \leqslant \phi \leqslant 45°\text{N}$, $70°\text{W} \leqslant \lambda \leqslant 80°\text{W}$ (M2 region).

The ET technique showed only small or moderate sensitivity for the M1 and M2 areas for all cases except for IOP11 and IOP13 (see Figs. 1(b) and 1(c)). In these two cases, M1 was highly sensitive and therefore, for the evaluation of the targeting algorithm based on the time evolution of the signal (section 5), conventional data from M1 were used as if they had been targeted observations in these two cases. However, when verifying the forecasts against analyses and observed data (section 6, Fig. 10 and Table 4), the data within the M1 and M2 regions were treated as regular non-FASTEX observations in all cases, including IOP11 and IOP13.

The amplification of the signal was defined in terms of the surface pressure p as

$$a(p, t_a, t_v) = \frac{\max_v |p_F(t_v) - p_c(t_v)|}{\max_a |p_F(t_a) - p_c(t_a)|}, \tag{1}$$

where the subscript F refers to the variable p in an experimental forecast with FASTEX targeted data and subscript c refers to the same variable in the control forecast. The analysis and the verification time are given in hours. The symbol $\max_a |\,.\,|$ ($\max_v |\,.\,|$) denotes the largest absolute value of p at a grid point within the targeted (verification) domain. Daily amplification factors (growth rates)

$$\delta(p, t_a, t_v) = a(p, t_a, t_v)^{24/(t_v - t_a)}, \tag{2}$$

were computed and compared for each IOP. The above expression offers a plausible norm that can be easily interpreted in the case of rapid atmospheric developments at the surface while it avoids the complexity of defining the proper weighting for different atmospheric variables in a more integrated norm.

In order to test the sensitivity of the results to the choice of error statistics in the analysis scheme, experiments were also carried out with the dropsonde error assumed in the data-assimilation scheme reduced for the dropsondes. The motivation for doing so is that in a region of rapid atmospheric development, the short-term forecast error is presumably larger than usual, thus the ratio between the errors in the background term and in the observations is larger than in an average case. In most of these experiments, the observation errors were reduced to 25 percent of the operational value, which means that the data received 16 times larger than normal weight. However, this strategy improves only the local fit of the analysis to the observed data; it will not improve the spatial structure of the balanced part of the analysis increment. This structure is determined by the standard background-error statistics. In the surface pressure, for instance, the balanced part of the analysis increment changes because of the better fit to the observed rotational wind component while the better fit to the observed surface pressure would alter only the analysis of the unbalanced component.

Finally, each forecast was verified against both the operational analysis and radiosonde observations that were available at the verification time.

TABLE 1. INVESTIGATED FASTEX CASES

IOP	Time targeting product issued	Analysis time	Verification time	Centre of verification region	Centre of most sensitive regions
9	0000 UTC 31 January	+24 h	+84 h	60.0°N, 15.0°W	50.0°N, 40.0°W
11	0000 UTC 03 February	+36 h	+72 h	52.5°N, 20.0°W	50.0°N, 35.0°W
					50.0°N, 60.0°W
13	0000 UTC 09 February	+36 h	+60 h	57.5°N, 20.0°W	50.0°N, 55.0°W
15	0000 UTC 13 February	+36 h	+72 h	50.0°N, 10.0°W	50.0°N, 40.0°W
					40.0°N, 50.0°W
16	0000 UTC 15 February	+36 h	+60 h	55.0°N, 10.0°W	40.0°N, 60.0°W
17	0000 UTC 17 February	+24 h	+48 h	52.5°N, 27.5°W	45.0°N, 50.0°W
					40.0°N, 65.0°W
18	0000 UTC 20 February	+48 h	+84 h	52.5°N, 17.5°W	50.0°N, 50.0°W

Observation and verification times are in terms of forecast lead time from time of sensitivity calculations. Sensitivity is based on real-time NCEP ensemble-based guidance.

5. TARGETING RESULTS

(*a*) *The Ensemble Technique sensitivity results*

Table 1 lists the most important parameters of the experiments: the targeting and verification times and locations. The coordinates of the most sensitive regions are based on the sensitivity maps shown in Fig. 1, which are identical to the real-time products that were used for making the flight decisions. All maps presented here are based on the identity-error covariance matrix, except for IOP9 in which case the diagonal estimate of the analysis-error covariance matrix from Iyengar *et al.* (1996) was used. Table 1 only lists those centres which could be continuously tracked across the Atlantic as the targeting time was brought closer to the verification time. Centres which failed to show this time continuity (e.g. the centre at 65°N, 20°W shown in Fig. 1(a)) were judged to be spurious in the real-time FASTEX decision making process. Since the ability of the ensemble transformations to represent the effect of changes in analysis-error variance on prediction-error variance was significantly constrained by the small size of the NCEP ensemble, we were not surprised to see such spurious correlations. Toth and Szunyogh (1997) found that the occurrence of such spurious centres could be almost completely eliminated by applying the ET method to a 50-member ensemble (available from ECMWF).

(*b*) *Observational and analysis increments with the targeted data*

The most obvious measure for quantifying the accuracy of a background field is the observational increment, the difference between the observations and the first-guess field. The main advantage of this verification is that errors in the observations and the first guess are uncorrelated and since the former is random by nature it does not contribute to the average increment. However, one should keep in mind that the observational increments are also affected by representativeness errors (due to measuring at a point instead of over an area or volume represented by the model), and interpolation errors. Therefore, all these error sources must be reflected in our estimate of the observational error. The observational increment can then be thought of as the difference between the observational error and the true error in the first-guess forecast.

As mentioned earlier, in the version of the SSI analysis scheme that was in operational use during FASTEX, the background field was not interpolated in time to observation times over the second half of the 6-hour assimilation period. In Fig. 2 we show the observational increment for surface-pressure observations from all 86 dropsondes used

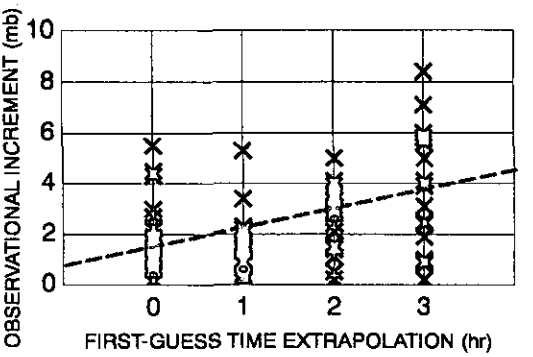

Figure 2. Surface-pressure observational increment as a function of length of extrapolation time of first-guess field to observation time. The linear-regression fit to the data ($r = 0.42$) is shown as a dashed line.

for the seven IOPs studied. It can be seen from Fig. 2 that the observational increments become larger as the 6-hour first-guess forecast is simply extrapolated out to a 3-hour lead time into the second half of the analysis window. The extrapolation introduces a substantial *additional* error that, in the case of a 3-hour extrapolation (2.0 hPa additional error), can be even larger than the 1.7 hPa error assumed for all surface-pressure observations. The average observational increment for the first 3 hours with the proper time interpolation is 1.3 hPa, while the overall average, including the extrapolated last 3 hours, is 2.3 hPa.

A large observational increment from a single dropsonde observation by itself does not necessarily trigger a large analysis increment (difference between the final analysis and the background forecast field). Other targeted dropsonde data, as well as satellite data and traditional observations over land, will also have an influence on the analysis increment. An inspection of the difference between the surface-pressure observational and analysis increments (difference between the observations and the analysis) reveals that the SSI analysis scheme performs rather well in statistical terms. The analysis, on average, draws as close to the observations as the observational errors indicate it should (approximately 1.7 hPa). One should note, however, that a good local fit to the data by itself does not guarantee that the structure of the analysis increment projects well onto the difference between the first guess and the true (unknown) state of the atmosphere.

Recall that the SSI, like other operational analysis schemes, uses nearly homogeneous and isotropic, background-error variance and covariance structures that are fixed in time. In other words, the analysis has no information regarding the likely, real-time background-error structures and magnitudes. Table 2 lists the maximum observational increment for the seven investigated IOPs. Not surprisingly, we found that in these sensitive areas the analysis typically remains twice as far from the observations as on average.

In the ET technique the temporal variation of the estimated analysis-error variance at t_a is estimated by the temporal variations of the ensemble-forecast spread. In order to verify whether this approach was valid for the FASTEX observations, we first have to verify whether the ensemble spread at future analysis times during FASTEX (36-hour forecast lead time) had a relationship with the first-guess errors. We show in Fig. 3 the relationship between the mean, surface-pressure observational increment and the corresponding NCEP ensemble spread for the seven investigated IOPs. The correlation between ensemble spread (at 36-hour lead time) and short-range (6-hour) forecast error on these spatial scales is 0.78, statistically significant at the 5% level. By averaging the

TABLE 2. THE MAXIMUM SURFACE-PRESSURE OBSERVATIONAL INCREMENTS FROM THE FASTEX DROPSONDE DATA AND THE MAXIMUM CHANGE CAUSED BY THESE DATA IN THE ANALYSIS AND THE FORECAST WITHIN THE *VERIFICATION REGION AT THE VERIFICATION TIME*

Intensive Observing Period	Observational increment (hPa)	Change in the analysis (hPa)	Change in the forecast (hPa)
9	1.1	0.5	2.2
11	–	1.2	2.2
13	4.5	1.2	1.3
15	5.4	3.0	3.6
16	1.4	0.7	1.0
17	4.3	1.4	1.2
18	2.2	1.0	2.0
Absolute average	3.2	1.3	1.9

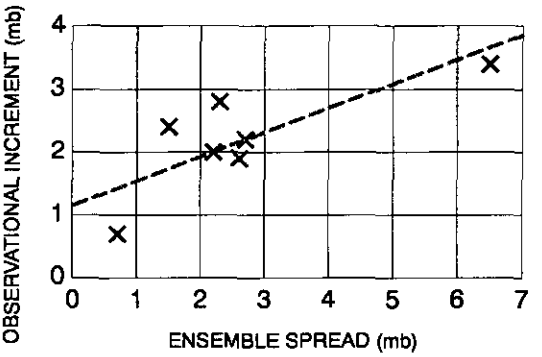

Figure 3. Surface-pressure observational increment as a function of ensemble spread at 36-hour forecast lead time for observations averaged for each of the seven IOPs (see text) investigated. The linear-regression fit to the data ($r = 0.78$) is shown as a dashed line.

errors for a total of 9–18 observations in each IOP case, the impact of a substantial part of the instrumental error and some of the representativeness error is eliminated. However, the time-extrapolation errors still affect the results. When the extrapolation factor is included in a bi-linear interpolation, the correlation between ensemble spread and short-range error (plus extrapolation error) increases to 0.84, explaining 70% of the variance. Notice also that the largest (smallest) observational increments occurred for the case when the ensemble spread was largest (smallest). In other words, Fig. 3 suggests that the 36-hour ensemble spread was a reasonable indicator of expected 6-hour forecast errors at a lead time of 36 hours for the FASTEX cases. However, owing to the small sample size the statistical significance of this result is delicate: the correlation may fall apart by removing one critical point. On the other hand we note that we obtained similarly high correlation values using the spread from the independent, 50-member ECMWF ensemble. While further research with a larger statistical sample is needed in order to confirm the relationship, indicated here, between ensemble spread and short-range forecast error, our result suggests that the 36-hour ensemble spread was a good indicator for the investigated forecast errors.

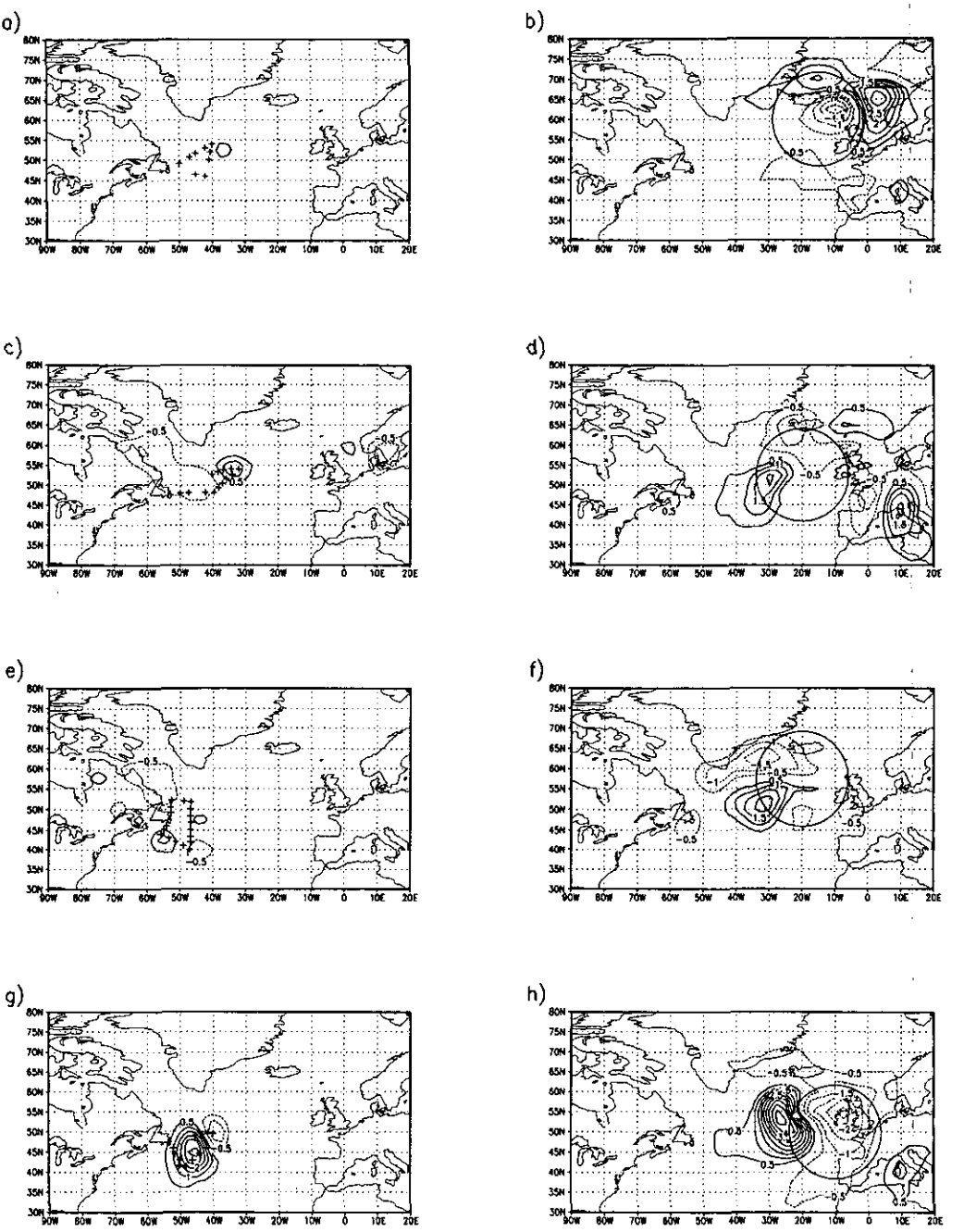

Figure 4. The differences between the surface-pressure analyses and ensuing forecasts with and without the Lear-jet dropsonde data for IOP9 ((a) and (b)), IOP11 ((c) and (d)), IOP13 ((e) and (f)), IOP15 ((g) and (h)), IOP16 ((i) and (j)), IOP17 ((k) and (l)), and IOP18 ((m) and (n)) (see text). Contour interval is 0.5 hPa.

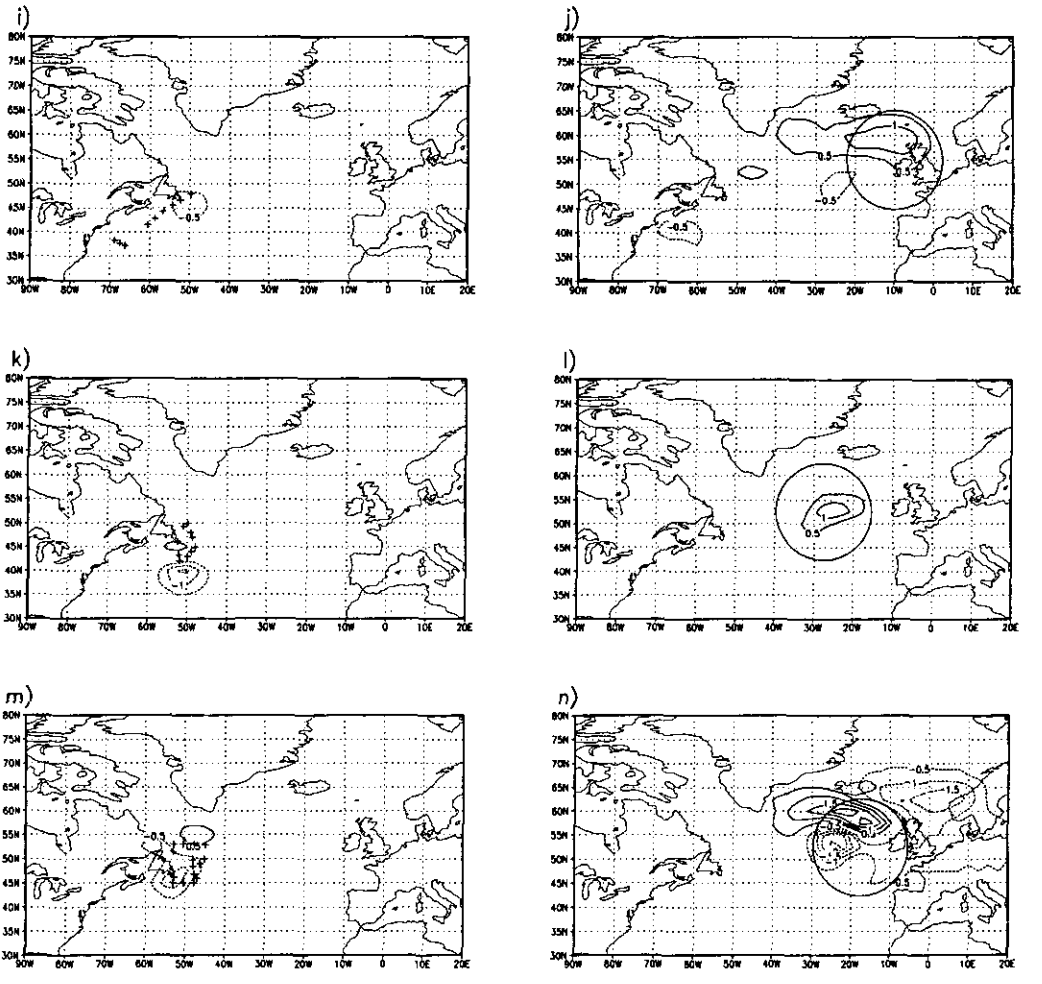

Figure 4. Continued.

(c) *Time evolution of the signal*

An important and necessary condition for a targeting methodology to be useful is that the initial signal created by the targeted data in the analysis (compared to using just the standard non-targeted observations) reaches the verification area. Shown in Fig. 4 are these analysis and forecast differences for surface pressure. Targeting with the ET method meets the necessary condition in the sense that a substantial portion of the dominant signal reaches the verification area in each of the seven FASTEX cases investigated here.

The absolute maximum amplitude of the signal at the analysis time and at the verification time within the verification region is given in Table 2. As the daily amplification factors indicate (Table 3), the growth of the signal is not uniform among the cases. For IOP18, the signal is decaying at the higher levels (not shown), though for the surface pressure the amplification rate is 1.8. For IOP17, on the other hand, there is no amplification at the surface, while for the 250 hPa height (not shown) the amplification rate is

TABLE 3. NORMALIZED DAILY AMPLIFICATION FACTORS FOR THE REAL AND THE MOCK (SEE TEXT) TARGETING REGIONS

Intensive Observing Period	Real targeting	Mock 1 experiment	Mock 2 experiment	ensemble
9	1.8	1.2	1.8	2.5
11	1.5	–	1.2	2.1
13	1.7	–	0.3	2.3
15	1.2	0.8	0.3	1.1
16	1.9	0.6	1.3	1.7
17	0.9	1.1	1.2	1.5
18	2.0	0.7	0.6	1.8
Average amplification	1.6	0.9	1.0	1.9

The amplification factor for the ensemble is based on the ratio of the maximum ensemble spread within the verification and targeting experiments.

5.0. As seen from Fig. 4, the eastward propagation of the signal in most cases is rather fast.

A general feature of the signal in these results is its small amplitude both at the targeting and verification times. Consequently, the time evolution of the signal is, to a good approximation, linear. This is demonstrated by an example from IOP13. The eastern part of the sensitive area for this IOP (cf. Fig. 1(c) and Table 1) was sampled by dropsondes released from the Lear jet over the ocean, while the western part (M1 area) was sampled by traditional radiosonde observations. Fig. 5 shows the signal at observation and verification times from the two parts of the sensitive area separately. If we add up Figs. 5(a) and 5(c), and 5(b) and 5(d), we recover, almost exactly, Figs. 4(e) and 4(f), respectively. Recall that Figs. 4(e) and 4(f) were produced by neglecting data from both the real and the M1 mock targeted region.

It is also interesting to note that the amplification averaged for all investigated cases for the first 6 hours is only 1.0. For IOP16, IOP17 and IOP18, for example, the signal decays initially. This suggests that in the beginning the changes in the analysis, due to the dropsonde data, have substantial projections on decaying and neutral structures that allow the signal to amplify only after an initial adjustment period. The signal for IOP17 experienced the largest decay (0.7/6 hours) and it started its rapid amplification just around the verification time. These two factors explain why one of the most dramatic IOP developments of FASTEX is associated with only a moderate growth rate in our calculations. Note that, at the time the Lear-jet flight was planned, the verification time listed in Table 1 (a total of 84 hours in the future) appeared to be the most likely time for an MSA flight that would survey the low-pressure system providing valuable data for verifying the targeted forecast.

Even though the small signal that the extra dropsonde data generates behaves linearly in time, the full-analysis increment does not necessarily do so. Note that the full-analysis increments obtained by using all the data have up to three times larger initial amplitudes than the signal from the dropsondes only. If we consider the time evolution of the error in the analysis and ensuing forecast, nonlinearity plays an even more important role. Since the perturbation size in the NCEP ensemble is of the order of forecast-error sizes (Zhu *et al.* 1996), the ensemble perturbations around the control offer a good representation of the uncertainty in the control analysis and forecast. While the performance of the ET technique is convincing in all of the investigated FASTEX cases, the fact that the signal has a much smaller amplitude than that of the

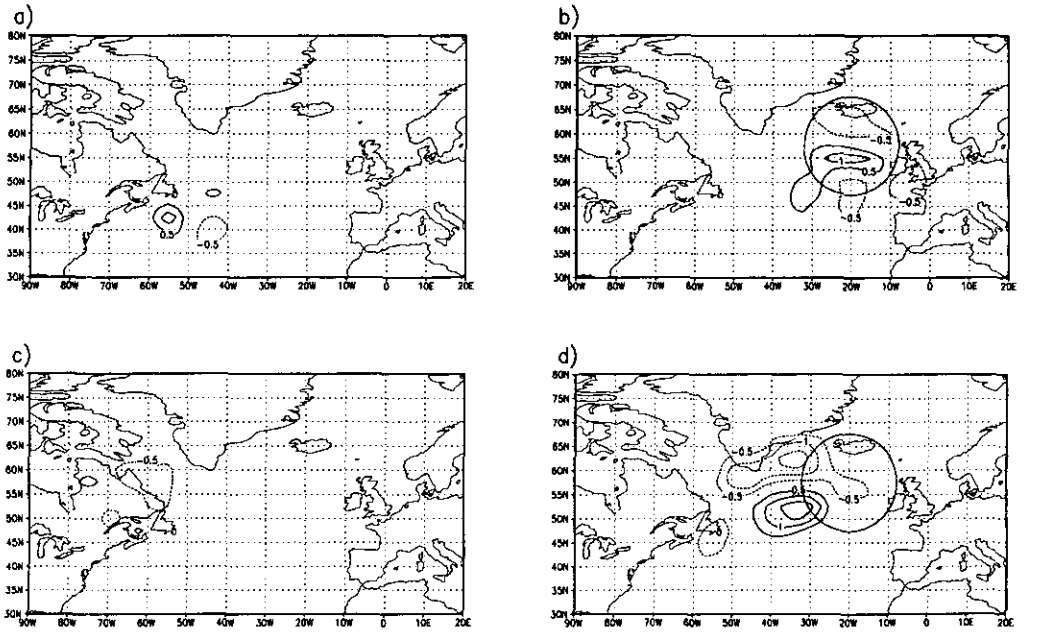

Figure 5. The differences between the surface-pressure analyses and ensuing forecasts with and without the Lear-jet dropsonde data ((a) and (b)) and with and without the regular data in the mock targeting region (M1) ((c) and (d)) for IOP13 (see text). Contour interval is 0.5 hPa.

forecast errors raises two questions. (1) What is the optimal perturbation size of an ensemble to be used for targeting? Our provisional answer is that the ensemble spread should be representative of the size of the full-analysis increments. The operational ensemble has larger perturbation amplitudes than this, corresponding to the total error in the analysis–forecast system. (2) Around what trajectory should an ensemble with such a smaller spread be centred? Given an uncertainty in the control forecast that is considerably larger in magnitude than the analysis increment, ideally we should carry out sensitivity calculations around each of the operational ensemble forecasts. A conservative approach would then be to mark all areas that are found sensitive in any of the sensitivity calculations, based on smaller-spread ensembles centred around each of the members of the larger-spread operational ensemble. Short of such an ideal system, one can reduce the perturbation amplitudes in the ET calculations based on the operational larger-spread ensemble. During NORPEX (NORth Pacific EXperiment)-98 and CALJET (California Landfalling Jets experiment) (Langland *et al.* 1998; Szunyogh *et al.* 1999; Ralph *et al.* 1999) we defined the perturbations as departures from the control forecast (rather than differences between positively- and negatively-perturbed forecasts, as used in this study). The use of an ensemble with more members would allow us to define even smaller perturbations as differences between selected members that evolve similarly.

(*d*) *Comparison of the real and the mock targeting results*

The fact that the targeted signal, as shown in the previous section, reaches the verification region is necessary but not sufficient to demonstrate that a targeting method is able to find the region that has the largest potential impact on forecast quality. Areas other than those targeted may have an even larger impact in the verification area.

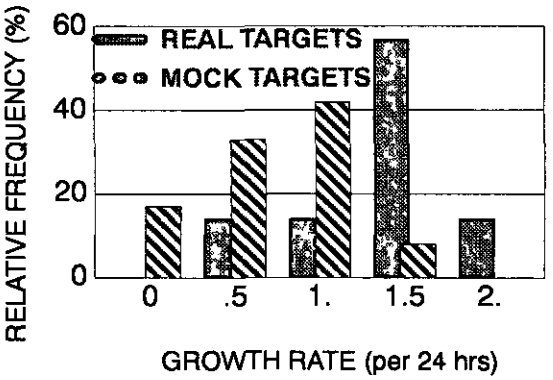

Figure 6. Relative frequency of growth rates for the real (solid bars) and mock (hatched bars) targeting signals for the seven IOPs (see text). Growth rates are grouped in intervals of 0.5, starting each group at the values indicated below the horizontal axis.

Whether this is true with the ET sensitivity results was tested through the mock targeting experiments described in section 4, by withholding regular data from areas adjacent to the real targeting region.

The size of the impact of targeted data at verification time in the verification region depends on (1) how large the effect of the data is at analysis time (initial signal), and (2) how the initial signal amplifies into the verification region. As to the size of the initial influence, it is on average 1.0 hPa for the mock, and 1.3 hPa for the real targeted signal. The smaller initial signal over the mock areas may be partly due to the generally better quality of the first guesses and analyses over these continental areas, and partly to the fewer number of regular radiosonde observations that could be eliminated from there, as compared to the oceanic real targeted areas.

The amplification factors for the real and mock signals are listed in Table 3. The average amplification for the real signals is 1.7, corresponding well with the amplification rates for bred perturbations with small amplitudes (Toth and Kalnay 1997). In contrast, the average amplification rate for mock targeting experiments is slightly less than 1.0, indicating, on average, a slight decay of the perturbations from the mock areas. Though signals from adjacent areas can still reach the verification region, the ET method is clearly capable of identifying areas of amplifying analysis errors. This is highlighted by the fact that the distributions of amplification factors for the real and mock signals are, according to a Wilcoxon rank sum test (Kendall and Stuart 1967), statistically different at the 0.1% level (Fig. 6).

When comparing the amplification factors from the continental mock areas with those from the real sensitive areas over the Atlantic Ocean we must keep in mind that the density and the total number of the regular radiosondes removed from the mock areas is much lower than that of typical targeted dropsondes. Therefore, the mock experimental analyses (and the differences between control and experimental analyses) are more influenced by the background forecasts and background-error statistics used in the analysis, and are therefore also more in balance, than the real targeted signals created by the dropsonde data. The average signal amplification for the first 6 hours for the mock areas is 1.2, while it is only 1.0 for the targeted areas, indicating that the initial adjustment process (decaying or neutral perturbation behaviour) is evident only for the targeted signal. For example, the only case when a mock signal had an amplification larger than the targeted signal (IOP17), the targeted signal, but not the mock signals,

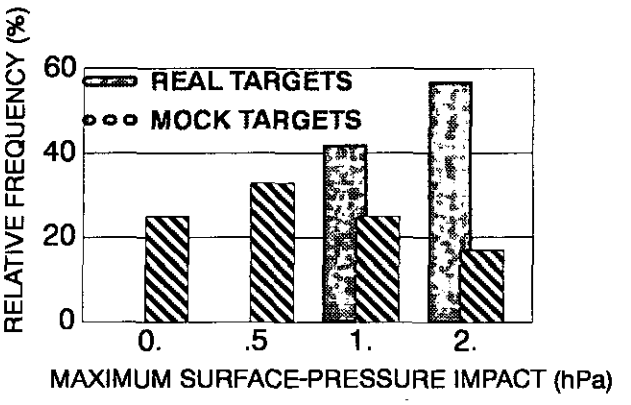

Figure 7. Relative frequency of maximum change in surface pressure from real (solid bars) and mock (hatched bars) targeted observations at verification time in the verification region. Values are grouped in intervals of 0.5, starting each group at the values indicated below the horizontal axis.

was affected by a strong initial adjustment process.* Disregarding the first 6 hours of signal evolution, the targeted signal had an amplification larger than the mock signals even for this IOP.

Though the amplification from the real sensitive areas is significantly larger, signals from the nearby areas usually reach the verification region as well. Interestingly, in some cases the signal from the mock areas adjacent to the most sensitive region has a structure in the verification area similar to that of the real targeting signal. In some sense, one may say that the instability that the targeted analysis/forecast uncertainty is associated with acts as a centre of 'gravity'. Perturbations from an initially large area can rotate in phase space (i.e. propagate in physical space) into the verification area at verification time.

When looking at the influence of the data at verification time in the verification region, the combined effect of the size of the initial signal and the amplification rate can be seen. While the average influence from the targeted areas is 1.9 hPa, that from the mock areas is only 1.0 hPa. The distribution of maximum impact values for the real and mock targeted areas is statistically different at a significance level lower than 0.5% (see Fig. 7).

The apparent skill the ET technique has in differentiating between more- or less-important areas of analysis uncertainty is due to the ability of the ensemble (1) to provide useful information, even at 36-hour lead times, regarding the expected background uncertainty, and (2) to approximate the dynamical propagation and amplification of a signal by how ensemble perturbations behave. We investigated the validity of this latter condition by comparing the amplification rates associated with the ensemble versus data signals. As the last column of Table 3 shows, the amplification factors calculated from the ensemble spread (i.e. ratio of maximum ensemble spread in the verification region at verification time, to the maximum ensemble spread within the targeted, most sensitive region) are generally higher than those for the signal. This is because (1) the ensemble perturbations are dynamically conditioned and are growing from the beginning of the sensitivity time period while the signal usually starts growing only after an initial adjustment period and (2) the ensemble amplification defined above is enlarged by the impact of perturbations from nearby areas (i.e. ensemble perturbations from areas not

* As it will be further discussed in section 6c, the analysis had obvious difficulties in assimilating data for IOP17.

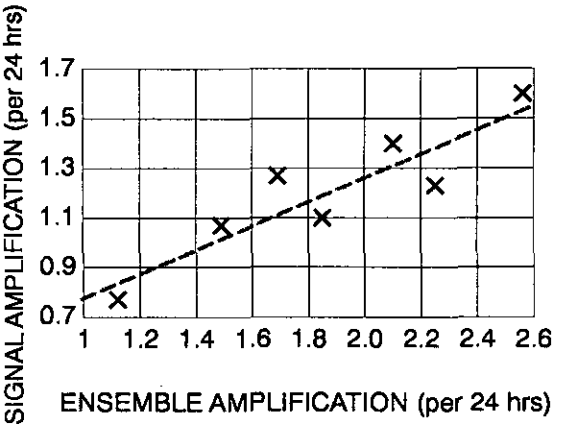

Figure 8. Amplification of the targeting signal (average for real and two adjacent mock areas) for the seven IOPs (see text) as a function of ensemble amplification. The linear-regression fit to the data ($r = 0.91$) is shown as a dashed line.

covered by the signal at the observation time can contribute to the size of the final ensemble perturbation in the verification region at the verification time).

Before comparing the signal and ensemble amplifications case by case, we recall that the targeted signal amplification can be seriously affected by an initial adjustment process (section 5c). Also, the mock signals appear to be associated with and project onto the same instability dominating the targeting signal. Therefore, to reduce sampling problems associated especially with the initial adjustment process, the signal amplification in Fig. 8 was determined for each case as the average of that from the real and two mock targeted areas. Fig. 8 seems to confirm a premise of the ET technique: a relationship exists between the time evolution of the signal associated with analysis increments and the perturbations in a bred ensemble.

6. Verification of forecasts

(*a*) *Verification against analyses*

The ultimate goal of targeting is to improve subsequent forecasts by assimilating the targeted data into future analyses. In this section we will compare the quality of the forecasts from analyses with and without the assimilation of the targeted data. Verifying the short-range forecasts against analysed fields has serious limitations. Recall that the average observational increment in surface pressure (which is a good indicator of the quality of atmospheric analyses because it reflects the performance of the analysis/short-range forecast system against independent data) for the seven IOPs in the upstream targeting domain is 2.3 hPa, with maximum errors in the dynamically most active regions being two to three times larger. The problems with errors in the verifying analyses are aggravated by the fact that the FASTEX verification region is on the eastern side of the Atlantic Ocean, with poor, upstream data coverage. Furthermore, the verification region and time are selected over areas of large forecast uncertainty, making the atmospheric analyses in these regions even more suspect.

In four cases (IOP11, IOP13, IOP16, and IOP17) the verification against the analysis showed a mixture of improvements and degradations at the different geographical locations (not shown). These results, however, are difficult to interpret because the amplitude of the signal in the verification region is of the order of the amplitude of the

TABLE 4. VERIFICATION AGAINST OBSERVED DATA

IOP	Verification region	Number of observations	Vector r.m.s. error without targeted data (m s^{-1})	Vector r.m.s. error with targeted data (m s^{-1})
9	55°N-70°N, 20°W-10°E	223	16.5	15.2
11	45°N-70°N, 30°W-00°E	378	9.0	9.0
13	45°N-65°N, 30°W-00°E	349	6.4	6.1
15	43°N-60°N, 35°W-00°E	256	6.4	7.5
16	50°N-65°N, 30°W-00°E	473	8.1	8.3
18	45°N-65°N, 30°W-00°E	294	10.2	9.9

For each Intensive Observing Period (IOP), errors are shown for the control forecast and the forecast with targeted data.

analysis errors (see Figs. 4(d), 4(f), 4(j), and 4(l)). For the other three cases, the absolute error in the surface-pressure, control forecast fields and the reduction in these errors due to the use of targeted data are shown in Fig. 9. Negative (positive) values in the right panels indicate forecast improvement (degradation) due to the inclusion of targeted data. The forecast with targeted data for IOP9 (Figs. 9(a) and (b)) is unambiguously improved: the timing of the cyclonic wave in the verification region is better, thus reducing a large dipole in the forecast error. The improvement in forecast quality (Fig. 9(c)) becomes even larger when the observational errors for dropsonde observations are reduced in the analysis by a factor of four and consequently the analysis can draw even closer to the targeted data. Interestingly, the observational increments at targeting time were by far the smallest for this IOP (see Table 2, left column). For IOP18 (Fig. 9(g)) the verification results are less clear but suggest an overall improvement in forecast quality. For IOP15 (Fig. 9(e)) the inclusion of the targeted data had a negative overall effect on the quality of the forecast. However, there is no considerable forecast error in the MSA for this case leaving virtually no room for forecast improvement.

(*b*) *Verification against observed data*

As mentioned earlier, most IOPs did not provide sounding data for the verification of the targeted forecasts. However, for all cases, except for IOP17 (where the signal is exclusively over the Atlantic Ocean), regular radiosonde observations are available for a comparative verification of the control and the targeted forecasts in and around the verification region. The scores discussed below have been computed for an extended region covering all areas with a sizable forecast-difference signal. The verification results, using all available radiosonde data between the surface and 250 hPa height, are shown in Table 4.* Note that while the total number of data involved in the verification is large, the data coverage is far from uniform. The areas of Iceland, the Norwegian Sea, the North Sea and Great Britain are much better sampled than the Atlantic Ocean.

Confirming the verification results of the previous section, IOP9 shows the largest forecast improvement in Table 4. The mean error (not shown) is reduced by 10–20%, while the r.m.s. error is reduced by close to 10%. Increasing the local fit of the analysis to the observed data further improved the forecast, (the vector r.m.s. error was 14.9 m s^{-1}), which is again in good agreement with verification results using the verifying analysis. IOP9 was clearly a success in the sense that the analysis with the targeted data yielded a forecast superior to the control.

* Due to a programming error, the verifications for IOP9, IOP13, and IOP15 included a few radiosonde observations that fell outside the verification boxes indicated in Table 4. It is most likely, however, that because of the large total number of observations these extra data have no measurable impact on the the corresponding verification results shown in Table 4.

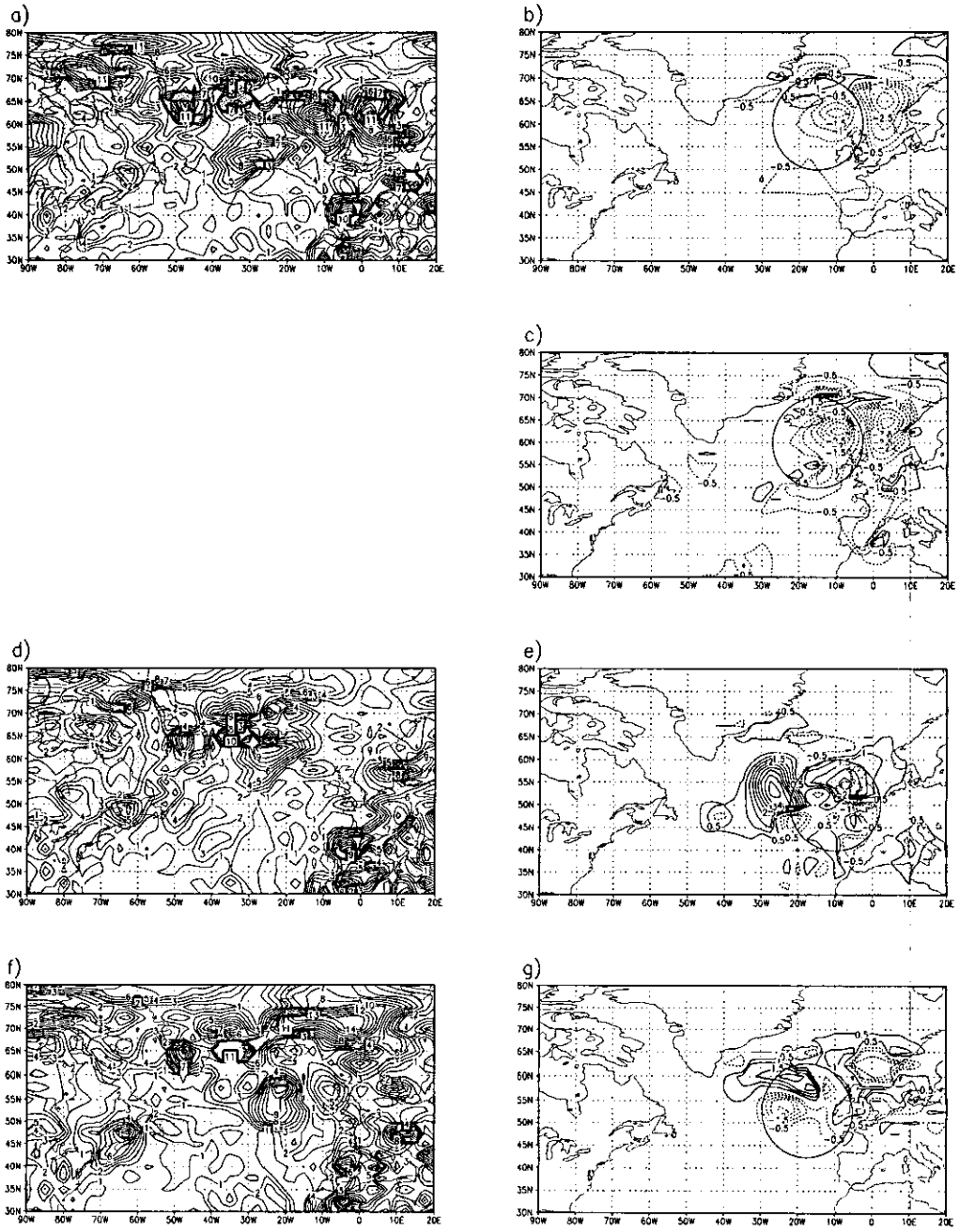

Figure 9. The absolute error in the control forecast (left panels, contour interval is 1.0 hPa) and the absolute forecast-error reduction in the surface pressure resulting from the Lear-jet dropsonde data (right panels, contour interval is 0.5 hPa) from the operational values of the observation errors for IOP9 ((a) and (b)), IOP15 ((d) and (e)), and IOP18 ((f) and (g)) (see text). Panel (c) shows the absolute forecast-error reduction using 75% smaller, estimated dropsonde observation errors than in the operational assimilation scheme for IOP9. Negative (positive) values in the right panels indicate forecast improvement (degradation) due to the inclusion of targeted data.

Targeted data assimilated for IOP18 produced the second largest reduction in forecast errors. It is interesting to note that the vector r.m.s. error was reduced even further (to 9.7 m s^{-1}) when the targeted data were assimilated with reduced observation errors. This is despite the somewhat mixed results when the analysis (and not observed data) was used for verification (see Fig. 9(g)). A similar improvement is seen in the case of IOP13, while the targeted data had no impact on forecast quality in the case of IOP11. The quality of the forecast decreased slightly for IOP16, while in the case of IOP15, corroborating the results from the previous section (see Fig. 9(e)), the inclusion of the targeted data clearly degraded the forecast. The above results are in qualitative agreement with those of Thorpe (1999), regarding the influence of including targeted dropsonde data in the ECMWF analysis/forecast system. Their initial results indicate that, similar to our findings, the effect of the data on forecast quality at short lead times is relatively small, and is generally positive, but in some cases the use of the extra dropsonde data slightly degrades the forecasts.

We note that the analysis is a statistical procedure and forecast improvements from the additional data are to be expected only in a statistical sense, and not in every case. Still, from a practical point of view, it is not comforting to know that in some cases the impact of extra data, produced at a considerable cost, may be negative. From this perspective it is important to point out that the forecast improvement the targeted data brought about in our experiments is correlated with the error in the control forecast. The larger (smaller) the control error, the more improvement (possible degradation) we can expect. Intuitively, one can easily understand that it may be hard or even impossible to reduce the error in a very good forecast while a poor forecast may more readily lend itself to improvements. This is the basis for the selection of FASTEX IOP cases for the study of Thorpe (1999) for example.

Forecast error, of course, can be known only retrospectively and therefore cannot provide the basis for operational targeting strategies. However, an ensemble of forecasts available in the real-time decision making process, as found in earlier sections, may provide potentially vital information regarding expected forecast-error behaviour. In fact it was found (see Fig. 10) that ensemble amplification (left column in Table 3) is strongly related to forecast improvement (difference between the two right-hand columns of Table 4). The linear correlation is 0.91 for the six IOPs considered here and is statistically significant at the 0.5% level. On the other hand there is no plausible explanation why neutral or decaying signals should consistently degrade the forecast. Independent data collected during the NORPEX-98 field experiment (Szunyogh *et al.* 1998) confirmed that a large ensemble amplification is a good indicator of forecast improvement, but for weakly amplifying or decaying ensemble signals the forecast can be either improved or degraded by the assimilation of extra dropsonde data. These results suggest that targeted data have a good chance of improving forecasts only if the ensemble spread from the most sensitive region at targeting time amplifies considerably into the verification region at verification time.

Note that ensemble amplification, found here to be a good indicator for the forecast improvement expected from targeted data, was considered in real time during FASTEX but was not the main factor in selecting targeting cases. Instead, IOPs were selected primarily with respect to the needs and expectations of the scientists responsible for downstream observations in the MSA.

(*c*) *Intensive Observing Period 17*

One of the most impressive and most extensively studied cyclone development cases from the FASTEX period is IOP17. As mentioned earlier, the signal from the dropsonde

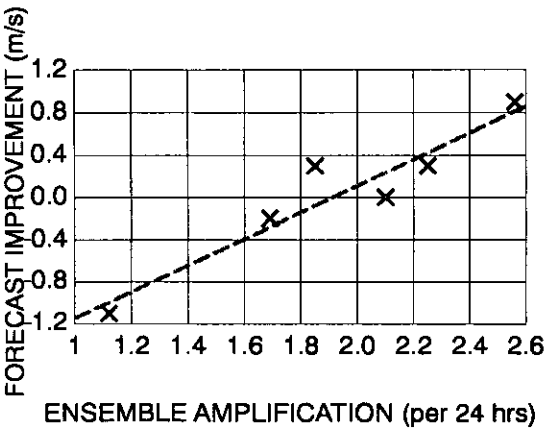

Figure 10. Forecast improvement (root mean square, wind-vector, error reduction) for six IOPs (see text) (with enough observations to allow the verification) as a function of ensemble amplification. The linear-regression fit to the data ($r = 0.91$) is shown as a dashed line.

data was situated over the data-void Atlantic at verification time. However, 6 hours after verification time, the UK C-130 collected valuable data which we used for verifying forecasts with lead times of 30 and 42 hours. Note that the dropsonde data were used in one of the two 30-hour forecasts. The observational errors were reduced compared with the operational set-up, in order to increase the amplitude of the signal by forcing the analysis closer to the data. Interestingly, the best forecast was the 42-hour control forecast with a small, 3 mb error at the centre of the low. Adding more and more data at later analysis times, including the special dropsonde data at 30-hour forecast lead times, only made the forecasts worse and worse: the 36-hour lead time error was 5 mb, the 30-hour control error was 8 mb, while the 30-hour error for the forecast using the dropsonde data was 11 mb.

The fact that the addition of data in three instances (at two assimilation times) made the quality of an otherwise good 42-hour forecast substantially poorer suggests that for IOP17 the statistical assumptions made in the analysis were insufficient for the successful assimilation of the observations. The verification against the C-130 dropsonde data confirmed the above conclusion: at 30-hour forecast lead times the r.m.s. wind-vector error is 13.7 m s^{1} (14.5 m s^{1}) without (with) assimilating the targeted data.

Note that IOP17 was associated with the second-lowest ensemble amplification. Based on the results shown in Fig. 10, the poor, targeted forecast performance could have been anticipated.

7. Conclusions

In this paper we evaluated the performance of the ensemble-based targeting methodology as used in real time during the FASTEX field experiments, based on the operational NCEP global ensemble forecasts. The results, including those of analysis–forecast experiments with and without the targeted data, demonstrate that the ensemble-based targeting technique as applied in real time during FASTEX is a useful tool in the following senses:

- The spread (standard deviation around the mean) in the operational ensemble is a reasonable indicator of the expected error variance in the first-guess short-range forecast fields used in the analysis procedure (see Fig. 3);
- The evolution of perturbations in the ensemble is a good indicator of how signals introduced by (targeted) data evolve in time (Figs. 4 and 8);

- The ET technique (Bishop and Toth 1998) can identify the area from which initial perturbations can have the largest influence on forecast quality in a pre-selected verification area at a given verification time. In particular, it was found that in all seven cases studied, the initial signal introduced into the analysis by the targeted data from the most sensitive area reaches the verification region at the verification time (Fig. 4), and signals from adjacent areas had significantly less contribution (Table 3, Figs. 6 and 7).
- There is an overall improvement in forecast quality from the analyses enhanced by targeted data, but only when the signal amplifies appreciably. The results suggest that, given the current observing/analysis/forecast schemes, we should attempt to take targeted observations only if the related signal, based on the evolution of the ensemble spread, is expected to amplify (Fig. 10). Otherwise, the errors inherent in the analysis/forecast system may hide the small signal resulting from the targeted data.

Though the above results are based only on those seven FASTEX cases from February in which the ET technique (Bishop and Toth 1998) was one of the primary guidance products used, they are all statistically significant, at least at the 5% or higher level. These results suggest that many important aspects of targeting may be addressed by the use of an objective targeting method, such as the ET technique, based on an operational ensemble of forecasts. A synthesis of the results yields the following provisional algorithm for targeted observations:

- Select a verification area and time, based on large forecast uncertainty estimated from the ensemble spread, also considering the socio-economic consequences of potential forecast errors;
- Consider an ensemble of forecasts that can assure an estimation of future analysis uncertainty and forecast-error development for the objective targeting method, such as the ET technique;
- Use the ET technique with the operational ensemble to identify the targeted region from which data can have the largest impact in the verification area;
- Consider how ensemble perturbations from the initial target area amplify into the verification area at verification time. Take observations only for cases of large amplification.

It is important to note that the influence of targeted data (9–18 dropsondes per case over an area of approximately 10 degrees of latitude by 10 degrees of longitude) on the forecasts is rather small. The maximum forecast-error reduction is on the order of 10%. Larger improvements can be expected only if a larger area is targeted, including a less sensitive but extensive ring around the most sensitive core area; and better analysis techniques can be used, with flow-dependent background-error variance and covariance estimates (Fischer *et al.* 1998). This would enable the analysis to extract considerably more useful information, that now remains hidden in the data.

The fact that targeting during FASTEX took place just off the coast of one of the best observed areas on earth may have also tempered the influence of targeted observations. In these experiments, the first-guess forecast fields must have had a higher quality than, say, those over the mid Pacific. In fact, preliminary results from targeting experiments, that took place over the north-east Pacific in the winter of 1997–98 (NORPEX-98 and CALJET), indicate that the targeted data had a greater positive impact on forecast verifying over the continental USA (Szunyogh *et al.* 1999).

So far we have assumed that all forecast errors are due to initial-value uncertainties. Though this may be a good first approximation, certainly there are errors that arise from model deficiencies. In these cases, forcing the initial condition closer to the data may not necessarily help improve the ensuing forecasts. How common and serious this potential problem is, and how it can be mitigated, should be a subject of future research

with analysis/forecast systems of different complexities. The influence of using higher-resolution analysis/forecast schemes should also be explored.

Regarding the targeting strategy used during FASTEX, the following improvements can be made.

- The most serious limitation of the ET method, as applied to the NCEP ensemble, is the potential presence of spurious, distant sensitivity maxima that may appear in addition to the actual most sensitive region (e.g. Fig. 1(a), centre at 65°N, 20°W). This is due to the inevitable random correlations present in any small set of ensemble forecasts. These spurious centres can be routinely removed subjectively, but for the implementation of a fully automated technique they must be eliminated objectively. Experiments with the use of the 50-member ECMWF ensemble (Toth and Szunyogh 1997) indicate that a modest increase in ensemble size (from 14 NCEP ensemble members) can almost completely eliminate the problem of spurious sensitive areas.
- The magnitude of perturbations considered within the ensemble used in the ET technique should preferably be of the order of the size of the analysis/forecast increments. These increments are definitely smaller than the perturbations used during FASTEX, defined as the difference between oppositely perturbed (negative–positive) pairs of forecasts. Instead, perturbations can be defined as difference fields between the control and any perturbed forecast.
- The sensitivity calculations are not yet fully compatible with the analysis system. For example, the influence of regular data during targeting is considered only indirectly and only in a gross sense. We plan to link better the sensitivity calculations with the analysis scheme, by introducing into the calculations information from the observational-error covariance matrix.

The first two of these improvements were implemented during the application of the ET targeting technique in the NORPEX-98 (Langland *et al.* 1998; Szunyogh *et al.* 1999) and CALJET (Ralph *et al.* 1999) field experiments in January–March 1998. According to a preliminary evaluation (Szunyogh *et al.* 1999) all major results of the current study have been confirmed by the NORPEX-98 and CALJET field programs.

ACKNOWLEDGEMENTS

We would like to thank the staff of NCEP, and in particular Dr Eugenia Kalnay and Dr Stephen Lord of the Environmental Modeling Center for their support of research and development in targeted observations that allowed NCEP to successfully participate in the FASTEX field experiments. Wan-Shu Wu kindly helped with setting up the analysis–forecast experiments and offered advice regarding the assimilation of the FASTEX dropsonde data with the SSI analysis system. We are grateful to David Burridge (director) and the staff of ECMWF who provided the ECMWF ensemble forecasts to be used in real time during the FASTEX field experiments. We would also like to thank the staff of the Research Aviation Facility of the National Center for Atmospheric Research, and the pilots and support team of Flight International Inc. for their superb conduct of the Lear-jet operations out of St. Johns and the rapid production of corrected dropsonde datasets. The critical comments of Milija Zupanski, Glenn White, Dave Parrish and one of the reviewers on an earlier version of the manuscript are greatly appreciated. We also thank Dr Stephen Nelson of the National Science Foundation and Paul Herzegh of NCAR for their extraordinary efforts to secure the Lear jet for the project at very short notice, owing to mechanical difficulties with the aircraft originally planned for the project. Part of the research was supported by the National Science Foundation under grant ATM-9634239.

REFERENCES

Barkmeijer, J., van Gijzen, M. and Bouttier, F. 1998 Singular vectors and estimates of the analysis error covariance metric. *Q. J. R. Meteorol. Soc.*, **124,** 1695–1713

Bergot, T. 1999 Adaptive observations during FASTEX: A systematic survey of upstream flights *Q. J. R. Meteorol. Soc.*, **125,** 3271–3298

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

Bergot, T., Malardel, S. and Joly, A. 1996 'Sensitivity and singular vectors calculation in the operational context of FASTEX'. Pp. 366–368 in Preprints of the 7th AMS Conference on Meso-scale Processes, 19–23 August 1996 Norfolk, VA, USA. American Meteorological Society

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

Bishop, C. H. and Toth, Z. 1996 'Using ensembles to identify observations likely to improve forecasts'. Pp. 72–74 in Preprints of the 11th AMS Conference on Numerical Weather Prediction, 19–23 August 1996, Norfolk, VA, USA. American Meteorological Society

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

Bollt, E. M. and Kostelich, E. J. 1998 Optimal targeting of chaos. *Phys. Let. A*, **245,** 399–406

Cherry, S. 1996 Singular value decomposition analysis and canonical correlation analysis. *J. Climate*, **9,** 2003–2009

Derber, J. C., Parrish, D. F. and Lord, S. J. 1991 The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, **6,** 538–547

Emanuel, K. A., Lorenz, E. N. and Morss, R. E. 1996 'Adaptive observations'. Pp. 67–69 in Preprints of the 11th AMS Conference on Numerical Weather Prediction, 19–23 August 1996, Norfolk, VA, USA. American Meteorological Society

Fischer, C., Joly, A. and Lalaurette, F. 1998 Error growth and Kalman filtering within an idealized baroclinic flow. *Tellus*, **50A,** 596–615

Gandin, L. S. 1990 Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Weather Rev.*, **118,** 2754–2767

Gelaro, R., Langland, R. H., Rohaly, G. D. and Rosmond, T. E. 1999 An assessment of the singular-vector approach to targeted observing using the FASTEX dataset. *Q. J. R. Meteorol. Soc.*, **125,** 3299–3327

Iyengar, G., Toth, Z., Kalnay, E. and Woolen, J. S. 1996 'Are the bred vectors representative of analysis errors?' Pp. 64–65 in Preprints of the 11th AMS Conference on Numerical Weather Prediction, 19–23 August 1996, Norfolk, VA, USA. American Meteorological Society

Joly, A., Jorgensen. D., Shapiro, M. A., Thorpe, A. J., Bessemoulin, P., Browning, K. A., Cammas, J. P., Chalon, J. P., Clough, S. A., Emanuel, K. A., Eymard, L., Gall R., Hildebrand, P. H., Langland, R. H., Lemaitre, Y., Lynch P., Moore, J. A., Persson, P. O. G., Snyder, C. and Wakimoto, R. M. 1997 The fronts and Atlantic storm-track experiment (FASTEX): Scientific objectives and experimental design. *Bull. Am. Meteorol. Soc.*, **78,** 1917–1940

Joly, A., Browning, K. A., Bessemoulin, P., Cammas, J.-P., Caniaux, G., Chalon, J.-P., Clough, S. A., Dirks, R., Emanuel, K. A., Eymard, L., Gall, R., Hewson, T. D., Hildebrand, P. H., Jorgensen, D., Lalaurette, F., Langland, R. H., Lemaître, Y., Mascart, P., Moore, J. A., Persson, P. O. G., Roux, F., Shapiro, M. A., Snyder, C., Toth, Z. and Wakimoto, R. M. 1999 Overview of the field phase of the Fronts and Atlantic Storm-Track EXperiment (FASTEX) project. *Q. J. R. Meteorol. Soc.* **125,** 3131–3163

Kalnay, E. and Pu, Z.-X. 1997 Using an adjoint model and/or a quasi-inverse linear model to target weather observations. *Research activities in atmospheric and oceanic modeling*. Ed. A. Staniforth. CAS/JSC/WGNE, WMO/TD-No. 792

Kanamitsu, M., Alpert, J., Campana, K., Caplan, P., Deaven, D., Iredell, M., Katz, B., Pan, H.-L., Sela, J. and White, G. H. 1991 Recent changes implemented into the global forecast system at NMC. *Weather and Forecasting*, **6**, 425–435

Kendall, M. G. and Stuart, A. 1967 Pp. 420–441 in *The advanced theory of statistics*, vol. 2. Hafner Publishing Co., New York, USA

Langland, R. H. and Rohaly, G. D. 1996 'Analysis error and adjoint sensitivity in prediction of a north Atlantic frontal cyclone'. Pp. 150–152 in Preprints of the 11th AMS Conference on Numerical Weather Prediction, 19–23 August 1996, Norfolk, VA, USA, American Meteorological Society

Langland, R. H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., Majumdar, S., Morss, R., Rohaly, G. D., Velden, C., Bond, N. and Bishop, C. 1998 The North Pacific-Experiment (NORPEX-98). Targeted observations for improved North American weather forecasts. *Bull. Am. Meteorol. Soc.*, in press

Lord, S. J. 1996 'The impact on synoptic-scale forecasts over the United States of dropwindsonde observations taken in the northeast Pacific'. Pp. 70–71 in Preprints of the 11th Conference on Numerical Weather Prediction, 19–23 August 1996, Norfolk, VA, USA. American Meteorological Society

Lorenz, E. N. 1990 'Effects of analysis and model errors on routine weather forecasts'. Vol. 1 Pp. 115–128 in Proceedings of ECMWF seminars on 'Ten years of medium-range weather forecasting', 4–8 September 1989, Reading, UK. ECMWF

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

Morss, R. E. 1999 'Adaptive observations: Idealized sampling strategies for improving numerical weather prediction'. PhD. Thesis, Massachusetts Institute of Technology

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

Parrish, D. F. and Derber, J. C. 1992 The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Weather Rev.*, **120**, 1747–1763

Parrish, D. F., Derber, J. C., Purser, R. J., Wu, W.-S. and Pu, Z.-X. 1997 The NCEP global analysis system: Recent improvements and future plans. *J. Meteorol. Soc. Jpn*, **75, 1B**, 359–365

Pu, Z.-X., Kalnay, E., Sela, J. and Szunyogh, I. 1997 Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Weather Rev.*, **125**, 2479–2503

Pu, Z.-X., Kalnay, E. and Toth, Z. 1998a 'Application of the quasi-inverse linear and adjoint NCEP global models to targeted observations during FASTEX'. Pp. 8–9 in Preprint of the 12th Conference on Numerical Weather Prediction, 11–16 January 1998, Phoenix, AZ, USA. American Meteorological Society

Pu, Z.-X., Lord, S. and Kalnay, E. 1998b Forecast sensitivity with dropwindsonde data and targeted observations. *Tellus*, **50A**, 391

Ralph, F. M., Persson, O., Reynolds. D., Nuss, W., Miller, D., Schmidt, J., Jorgensen, D., Wilczak, J., Neiman, P., Bao, J.-W., Kingsmill, D., Toth, Z., Veldon, C., White, A., King, C. and Wurman, J. 1999 The California land-falling jets experiment (CALJET): Objectives and design of a coastal atmosphere–ocean observing system deployed during a strong El Niño'. Pp. 78–81 in Proceedings of the 3rd Symposium on Integrated Observing Systems, 10–15 January 1999, Dallas, Texas, USA. American Meteorological Society

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

Szunyogh, I., Toth, Z., Majumdar, S., Morss, R., Bishop, C. and Lord, S. 1999 'Ensemble-based targeted observations during NORPEX'. Pp. 74–77 in Proceedings of the Third Symposium on Integrated Observing Systems, 10–15 January 1999, Dallas, Texas, USA. American Meteorological Society

Thorpe, A. J. 1999 'Predictability studies with FASTEX data using the ECMWF IFS'. Pp. 185–194 in Proceedings of the ECMWF workshop on predictability, 20–22 October 1997, Reading, UK, ECMWF

Toth, Z. and Kalnay, E. 1997 Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125**, 3297–3319

Toth, Z. and Szunyogh, I. 1997 'Review of the use of ECMWF ensemble data for targeting upstream observations during the FASTEX field experiment'. P. 131 in *Proceedings of the FASTEX Upstream Observations Workshop*, 10–11 April 1997, NCEP, Camp Springs, MD, USA. NOAA/NWS/NCEP Office Note 420. Environmental Modeling Center, 5200 Auth Rd., WWB, Rm. 207, Camp Springs, MD 20746, USA

Vanneste, J., P., Morisson, J. and Warn, T. 1998 Strong echo effect and nonlinear transient growth in shear flows. *Physics of Fluids*, **10**, 1398–1404

Woolen, J. S. 1991 'New NMC operational OI quality control'. Pp. 24–27 in reprints of the *9th AMS Conference on Numerical Weather Prediction*, 14–18 October 1991, Denver, CO, USA

Zhu, Y, Iyengar, G., Toth, Z., Tracton, M. S. and Marchok, T. 1996 'Objective evaluation of the NCEP global ensemble forecasting system'. Pp. 79–82 in Preprints of the 15th AMS Conference on Weather Analysis and Forecasting, 19–23 August 1996, Norfolk, Virginia, USA