

## The Effect of Targeted Dropsonde Observations during the 1999 Winter Storm Reconnaissance Program

I. SZUNYOGH\*

*UCAR Visiting Scientist, National Centers for Environmental Prediction, Camp Springs, Maryland*

Z. TOTH

*General Sciences Corporation, National Centers for Environmental Prediction, Camp Springs, Maryland*

R. E. MORSS

*Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts*

S. J. MAJUMDAR, B. J. ETHERTON, AND C. H. BISHOP

*Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania*

(Manuscript received 24 August 1999, in final form 6 March 2000)

### ABSTRACT

In this paper, the effects of targeted dropsonde observations on operational global numerical weather analyses and forecasts made at the National Centers for Environmental Prediction (NCEP) are evaluated. The data were collected during the 1999 Winter Storm Reconnaissance field program at locations that were found optimal by the ensemble transform technique for reducing specific forecast errors over the continental United States and Alaska. Two parallel analysis–forecast cycles are compared; one assimilates all operationally available data including those from the targeted dropsondes, whereas the other is identical except that it excludes all dropsonde data collected during the program.

It was found that large analysis errors appear in areas of intense baroclinic energy conversion over the northeast Pacific and are strongly associated with errors in the first-guess field. The “signal,” defined by the difference between analysis–forecast cycles with and without the dropsonde data, propagates at an average speed of 30° per day along the storm track to the east. Hovmöller diagrams and eddy statistics suggest that downstream development plays a significant role in spreading out the effect of the dropsondes in space and time. On average, the largest rms surface pressure errors are reduced by 10%–20% associated with the eastward-propagating leading edge of the signal. The dropsonde data seem to be more effective in reducing forecast errors when zonal flow prevails over the eastern Pacific. Results from combined verification statistics (based on surface pressure, tropospheric winds, and precipitation amount) indicate that the dropsonde data improved the forecasts in 18 of the 25 targeted cases, while the impact was negative (neutral) in only 5 (2) cases.

### 1. Introduction

The first quasi-operational Winter Storm Reconnaissance Program (WSR99) was carried out over the northeast Pacific between 13 January and 10 February 1999. Dropsondes were released in an effort to improve initial

conditions for short- and medium-range numerical weather forecasts over the continental United States and Alaska. The time and location of the dropsonde missions were chosen adaptively with the primary aim being to reduce the uncertainty in the prediction of land-falling winter storms and other weather systems with large potential societal impact.

Results accumulated using the operational analysis–forecast system of the National Centers for Environmental Prediction (NCEP) during the field programs of the North Pacific Experiment [NORPEX-98, January–February 1998; Langland et al. (1999); Szunyogh et al. (1999b)], and the California Land-falling Jets Experiment [CALJET, January–February 1998; Ralph et al. (1999); Szunyogh et al. (1999b); Toth et al. (2000)]

---

\* Permanent affiliation: Department of Meteorology, Eotvos Lorand University, Budapest, Hungary.

---

*Corresponding author address:* Dr. Istvan Szunyogh, National Centers for Environmental Prediction, EMC, 5200 Auth Road, WWB, Room 207, Camp Springs, MD 20746.  
E-mail: Istvan.Szunyogh@noaa.gov

showed that high quality dropsonde data collected in the upstream region of the Pacific storm track can make significant changes in the analyses, which led, in most of the cases, to an improvement of the quality of the ensuing forecasts over the United States.

These results based on field programs are in line with those for numerical experiments carried out with simplified low-order models of the atmosphere. As Emanuel et al. (1996), Lorenz and Emanuel (1998), Morss (1999), and Morss et al. (2000) demonstrated, adaptively taken observations have potential to improve weather analyses and forecasts when used in addition to traditional observations taken at fixed locations or as opportunities arise. Targeted observations over oceans can be collected by rawinsondes dropped from aircraft, unmanned aircraft, or satellite-based instruments. Added observations over the northeast Pacific are expected to improve the analysis and forecast of fast-moving storms that can cause heavy precipitation events in Alaska and the western states of the continental United States.

To explain how the effect of data propagates from the northeast Pacific into regions as far east as Europe, one has to call upon the concept of downstream development. In the conceptual model of storm tracks (Chang and Orlanski 1993) geopotential fluxes from a mature system on the western edge of the storm track propagate eastward in the upper troposphere, triggering new downstream baroclinic developments in regions of strong baroclinicity in the lower troposphere, and building upper-level downstream disturbances in areas of weak baroclinicity (Orlanski and Sheldon 1993). The downstream developments then generate new geopotential fluxes through baroclinic energy conversion, thus extending the storm track farther downstream. Analyzing data from seven winter seasons, Chang (1993) showed that downstream development can play an important role in the maintenance of the Pacific and the Atlantic storm tracks. In a thought-provoking study, Orlanski and Sheldon (1995) demonstrated, using a local energetics approach, that geopotential fluxes from a mature cyclone over the Pacific could trigger a chain of events that eventually led to the explosive baroclinic downstream development of the "Blizzard of '93." They argued that this result demonstrated that downstream development was a robust explanation for baroclinic development, in general, and conjectured that the improved analysis and forecast of the eddy packets and the energy conversion processes may improve the predictability of midlatitude storms, even beyond the life span of the individual systems.

In general, the successful operational implementation of an efficient targeting technique requires a decision-making procedure that is able to address the following issues in real time.

- *Case selection:* Identification of a weather event at *planning time*,  $t_p$ , for targeting at a *future observation time*,  $t_o$ . A weather event is selected, if it imposes a

major threat to the society at a later forecast *verification time*,  $t_v$ , and if it is judged that the uncertainty in this forecast can be considerably reduced by taking extra observations at the future observation time,  $t_o$ . Notice that  $t_p < t_o < t_v$ ; for instance, the time levels for the first WSR-99 flight were  $t_p = 0000$  UTC 11 January  $< t_o = 0000$  UTC 13 January  $< t_v = 1200$  UTC 14 January.

- *Sensitivity analysis and flight planning:* Identification of the most sensitive upstream area from where extra observations taken at  $t_o$  can most efficiently reduce the uncertainty in the prediction of the downstream weather event at  $t_v$ . Adaptive observational resources are added to sample efficiently the area identified by the sensitivity analysis. In our application, this accounts for the design of a flight track along which dropsondes can be released at designated locations.

The dropsonde data are then assimilated into an NWP model used to create short- and medium-range forecasts. For the first time, during WSR99, all components of the targeting procedure were treated in a quasi-operational manner. In this paper we evaluate the analysis-forecast performance of a targeting procedure that we suggest implementing operationally. Our study is based on the inclusion of data from 19 flights on 15 separate days, the largest sample available yet for a statistical evaluation of the impact of dropsonde data that were collected using the ensemble based targeting strategy.

The outline of the paper is as follows. Section 2 explains how the different components of the targeting procedure can be treated, emphasizing changes we made in our technique in order to support the operation-oriented application during WSR99. Section 3 briefly summarizes the WSR99 flight missions and describes the analysis-forecast system that was used to assess the value of the targeted data in improving numerical weather predictions. In section 4 we analyze the flow regimes that characterized the atmospheric circulation over the northeast Pacific during WSR99. In section 5, with the help of this analysis, we explore how the impact of dropsondes propagates in time. Section 6 presents forecast verification results, while section 7 offers some conclusions.

## 2. The targeting procedure

### a. Case selection

During the Fronts and Atlantic Storm-Track Experiment (FASTEX), the first experiment when targeting was tested in the field, there was little freedom in case selection for upstream observations. FASTEX was a multipurpose field experiment (Joly et al. 1997, 1999) and event selection was primarily based on the needs of scientists who designed downstream research flights in order to gain a better understanding of the dynamics of mature storms off-shore of Ireland. Likewise during CALJET, the main concern was to study the role of low-

level jets in land-falling heavy precipitation events over California (Ralph et al. 1999).

NORPEX (Langland et al. 1999; Szunyogh et al. 1999b) represented the first field experiment in which event selection was an integral part of the targeting procedure. This multiagency field program involved NCEP and the Naval Research Laboratory, Monterey, California. For each potential targeting scenario at NCEP, a verification region and associated verification time  $t_v$  were selected. These choices were primarily based on the time and location of the maximum in 24-h accumulated precipitation predicted over the United States at the planning time  $t_p$ . Another important criterion for selecting a case was the presence of considerable forecast uncertainty in the verification region, indicated by large spread in the operational NCEP ensemble forecast (Toth and Kalnay 1997) initiated at  $t_p$ . Quantitative probabilistic precipitation forecasts (Zhu et al. 1998) based on the same ensemble were also considered.

A common element in all prior field experiments was that the case selection was always done by the participating scientists. During WSR99, for the first time, the *weather events* were selected by the operational forecasters at the Hydrometeorological Prediction Center (HPC) of NCEP, after consultations with National Weather Service (NWS) field offices and other NCEP service centers. The exact verification time,  $t_v$ , and the center of a 1000-km-radius verification region (see Table 1) were chosen *in real time* by the authors of the present study following strategies similar to those used during NORPEX.

*b. Sensitivity analysis and flight planning*

*Sensitivity analysis techniques:* In recent years several different sensitivity analysis techniques have been suggested and tested during field experiments. Some of them are based on tangent-linear model integrations (e.g., Bergot et al. 1999; Gelaro et al. 1999; Palmer et al. 1998; Pu et al. 1997, 1998; Buizza and Montani 1999), while others are based on synoptic experience (Lord 1996) or on tracking potential vorticity anomalies. The latest technique was advocated by a group of scientists prior to FASTEX citing an influential paper by Hoskins et al. (1985), but to our best knowledge no results have been published on the evaluation of the performance of this technique during field programs.

The primary targeting technique at NCEP, which was used during WSR99, is based on an approach suggested first by Bishop and Toth (1996). This method, called the ensemble transform (ET) technique (Bishop and Toth 1999) can find linear statistical relationships between members of an ensemble of full nonlinear forecasts at  $t_o$  and the ensemble spread at  $t_v$ . During WSR99, ET calculations were performed using operational ensembles from NCEP (Toth and Kalnay 1997) and the European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza et al. 1998). The NCEP (ECMWF)

TABLE 1. Forecast error reduction in the verification regions. Here, W, E, and A, respectively, denote West Coast, East Coast, and Alaskan verification regions. Numbers in the same column show the verifying forecast lead times. The second and the third columns present the error reduction in percentages for the surface pressure,  $p_s$ , and the wind speed,  $|v|$ . In the last two columns  $+(-)$  means forecast improvement (degradation) for the precipitation (listed under Prcp) and the overall forecast skill (based on the previous three columns); 0 indicates neutral impact.

Date	Verif.	$p_s$	$ v $	Prcp	Summary
13 Jan 1999	W36	0.0	-1.0	+	0
17 Jan 1999	W24	10.7	0.0	0	+
	E60	12.8	14.0	0	+
19 Jan 1999	W36	12.0	5.4	+	+
	E84	10.4	7.4	+	+
20 Jan 1999	W24	-10.0	6.7	+	+
	E60	-32.1	-3.7	+	-
22 Jan 1999	W36	0.0	7.2	0	+
	E48	0.0	2.1	+	+
	A60	8.0	5.0		+
23 Jan 1999	W24	-4.0	1.5	-	-
	A24	7.1	2.4		+
27 Jan 1999	W48	12.5	-8.6	-	-
	E72	-6.3	-6.6	+	-
28 Jan 1999	W60	6.4	3.6	+	+
29 Jan 1999	W24	-12.5	-1.3	+	-
	E48	6.3	-1.3	+	+
3 Feb 1999	W36	10.0	8.0	0	+
	E48	8.7	1.1	0	+
5 Feb 1999	W36	9.1	0.0	0	+
6 Feb 1999	W24	9.1	1.2	+	+
7 Feb 1999	W24	9.5	-3.1	0	0
9 Feb 1999	E72	0.0	13.9	+	+
10 Feb 1999	W48	33.3	14.7	0	+
	E48	4.8	7.2	0	+
Wins/losses		17/4	16/7	12/2	18/5

ensemble consisted of seven (25) linearly independent T62 (T159) horizontal and 28 (31) level vertical resolution forecast members.

Here, we review only those aspects of the ET algorithm that are crucial to understanding the specifics of the WSR99 application. A more detailed description of this technique including its potential limitations and relationship with the adjoint-based methods is given in Szunyogh et al. (1999a). The same paper presents the results of “mock” targeting experiments, in which regular (nontargeted) data were withheld from areas adjacent to the real targeting regions that were selected based on the ET results. These results showed that for the FASTEX cases the technique could reliably distinguish between the areas that had the greatest contribution to the quality of the selected downstream forecast features. Similar results, based on cases in which dropsonde data were collected in regions that the ET technique found only moderately sensitive, are presented in Toth et al. (2000).

*General description of the ET technique:* The ET approach uses a set of ensemble perturbations  $\mathbf{x}_k(t)$  ( $k = 1, \dots, K$ ) defined as the difference between the  $K$  ensemble members and the ensemble mean (or the control forecast) at the future observation ( $t_o$ ) and verification

( $t_v$ ) times. The algorithm is based on the assumptions that 1) the ensemble perturbations span the space of the most likely deviations of the analysis–forecast from the truth at time levels  $t_o$  and  $t_v$ , and 2) the linear combinations of the ensemble perturbations will also represent likely deviations from the truth.

First, a set of verification variables is selected and a linear combination  $\mathbf{y}(t) = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \dots + c_K\mathbf{x}_K(t)$  is searched for, which maximizes a given distance function (norm)  $\|\mathbf{L}\mathbf{y}(t_v)\|^2 = \langle \mathbf{L}\mathbf{y}(t_v), \mathbf{L}\mathbf{y}(t_v) \rangle$  that measures likely deviation of the verification variables from the truth in the verification region at  $t_v$ . Here,  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  are the Euclidean scalar product and the associated norm for grid points that span the northeast Pacific including the verification region;  $\mathbf{L}$  is the localization operator that sets  $\mathbf{y}(t_v)$  to zero in regions outside of the verification area. The global magnitude,

$$\langle \mathbf{y}(t_o), \mathbf{A}^{-1}\mathbf{y}(t_o) \rangle = \sum_{i=1}^N \frac{1}{a_i} y_i^2(t_o), \quad (1)$$

of the linearly combined ensemble perturbations at the future analysis time,  $t_o$ , is chosen to be representative of the average of the global analysis uncertainty and kept constant throughout the maximization procedure. Here the  $a_i$  entries of the diagonal matrix  $\mathbf{A}$  denote the a priori estimate of the analysis uncertainty (the average rms distance between analyses from two independent analysis cycles) at the  $i$ th model grid point,  $y_i$  is the linearly combined perturbation at the same grid point, and  $N$  is the total number of grid points. In general,  $x_i(t_o)$  [and so  $y_i(t_o)$ ] is a vector whose components represent different hypothetical analysis variables at the same grid point  $i$ . We chose to use the wind at levels 850, 500, and 250 hPa as hypothetical analysis variables during the field program and the same variables were also used at verification time. The weights  $a_i^{-1}$  ( $i = 1, \dots, N$ ) allow for larger (smaller) deviations at  $t_o$  in regions where the analysis uncertainty is larger (smaller).

*Adaptive observations:* The influence of the hypothetical extra observations is introduced through reducing the a priori estimates of the analysis uncertainty  $a_i$  at the locations of hypothetical extra observations. At these locations the reduced uncertainty values will simulate the reduced analysis errors expected due to the hypothetical targeted observations taken at observation time  $t_o$ . The maximization is then repeated for each possible adaptation of the observational resources. The largest likely deviation of the forecast from the truth,  $G = \max\|\mathbf{L}\mathbf{y}(t_v)\|^2$ , will be reduced in every case when there is nonzero correlation between the ensemble perturbations in the region sampled by the hypothetical observations at the future analysis time  $t_o$  and within the verification region at verification time  $t_v$ . The optimal deployment of the observational resources is the one that most efficiently reduces the threat of a large forecast failure within the verification region, or formally, the one that most reduces  $G$ .

During the FASTEX, NORPEX, and WSR99 field programs the hypothetical targeted region was a 1000-km-radius disk centered at a geographical location  $\lambda, \phi$ , assuming that extra observations would reduce the analysis uncertainty by the same factor at each grid point within this region. The maximization was repeated many times, scanning a larger area that was available for targeting, at a  $5^\circ$  geographical resolution (the resolution of the targeting products should not be confused with the much higher resolution of the two ensemble forecasts used for the computations). A composite map displaying the value of  $G$  at the central location of the overlapping disks was then produced. Examples of the utilization of these maps for the individual FASTEX cases are shown in Szunyogh et al. (1999a), an example for NORPEX is presented in Langland et al. (1999), while the average of the sensitivity maps for WSR99 is shown in Toth et al. (1999). The preparation of the flight plan was based on such maps, with the aim of achieving the best data coverage in regions where the lowest values of  $G$  occurred.

For the first time, during WSR-99, the flights from Anchorage followed flight tracks that were designed before the field program started and the dropsonde locations were evenly distributed around these tracks. There were 20 of these flight tracks originally and 2 more were added later during the experiments. In addition to the traditional  $5^\circ$  resolution sensitivity maps, a new type of product was tested in the second half of WSR99, in which the maximization was repeated only 22 times, once for each predesigned flight track, instead of scanning a large geographical area with the overlapping disks. For these computations an enhanced version of the ET technique, the ensemble transform Kalman filter (ETKF; Bishop et al. 2000, manuscript submitted to *Mon. Wea. Rev.*) was used. In these computations the a priori analysis uncertainty was reduced at the fixed dropsonde locations along the flight tracks by assimilating hypothetical dropsonde observations using an ensemble Kalman filter. The optimal choice was the particular flight track that led to the largest reduction in the predicted forecast error variance within the verification region. This approach takes the impact of the future data into account in a more realistic way than the original ET technique and with the help of this method the second step of the targeting procedure (sensitivity analysis and flight planning) can be fully automated. Once the leading forecaster, who is responsible for the event selection, makes a decision regarding the verification time and the center of the verification region, that information can be used as input in the ETKF technique that can return the serial number of the optimal flight track. When the ET and the ETKF results were both available, the flight track was chosen by the authors subjectively weighting the two products. The detailed evaluation of the ETKF technique, which was done after the submission of the original version of the present paper, is beyond the scope of this study and will be the

subject of a forthcoming publication. We note, however, that based on the positive experience, we decided to use the ETKF technique with a combined ECMWF–NCEP ensemble (which consisted of 32 linearly independent members) during the operational Winter Storm Reconnaissance 2000 field program.

### c. Data analysis

The successful identification and sampling of an upstream region that is expected to have substantial influence on the forecast at  $t_v$  within the verification region does not guarantee that targeted observations will reduce the error in the prediction of the targeted event or even in a global sense. In fact, numerical experiments with low-order models of the atmosphere (Morss 1999; Morss 2000, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*) and experience accumulated during the field experiments (Szunyogh et al. 1999a,b) suggest that there is a nonnegligible risk that in certain cases extra observations can degrade, especially the longer than 2 day, forecasts. This is not completely unexpected. First, in a nonlinear system there is no guarantee that improvements in the initial conditions for a very limited number of components of the state vector will result in proportional improvements in a prediction. Second, the design of an operational analysis scheme that could completely remove the growing part of the analysis error is also impossible due to the chaotic nature of the atmosphere and the errors in the observed data (Pires et al. 1996). Third, the operational data assimilation procedures are statistical schemes that can guarantee the positive influence of the data only in a statistical time-average sense. On the practical side, there is a class of synoptic-scale events that are sensitive to the initial conditions but still well predictable on the short time range. Whether the data can improve the initial conditions in a way that leads to improvements in the forecasts of these systems depends on the ability of the analysis system to extract useful information from the observed data. Studies with low-order models of the atmosphere (e.g., Fischer et al. 1998) demonstrated that the assimilation of targeted data can be extremely sensitive to formulation problems in the analysis schemes. Many of these formulation problems are related to assumptions that facilitate working with large matrices in the operational practice and cannot be easily improved. But, there are some assumptions that can be relaxed and in the next section we will point to those recent changes in the NCEP analysis system that may contribute to the growing success of our targeting efforts.

## 3. The creation of the WSR99 dataset

### a. The flight missions

During WSR99, close to 500 dropsondes were released from 19 flights on 15 separate days. Eleven mis-

sions were carried out by one single plane, while on four occasions two planes were used for sampling selected areas over the Pacific. The missions involved 9 flights with two United States Air Force (USAF) Reserve C-130 planes flying out of Anchorage, Alaska, and 10 flights with the Gulfstream G-IV jet, operated by the National Oceanic and Atmospheric Administration (NOAA), and stationed in Honolulu, Hawaii. All C-130 flights originated and recovered in Anchorage, and most G-IV flights in Honolulu. One G-IV flight (0000 UTC 3 Feb; see Table 1 and Fig. 4), however, recovered on the U.S. mainland (Spokane, WA), while the two subsequent flights recovered in Anchorage and Honolulu, respectively.

### b. The analysis–forecast system

At NCEP, global forecast–analysis cycles were run operationally at T126 and T62 horizontal resolutions during WSR99. The dropsonde data were assimilated, in real time, into all operational analysis cycles by the spectral statistical interpolation (SSI; Parrish and Derber 1992; Parrish et al. 1997), a 3D variational assimilation scheme. All dropsonde data were quality controlled by the optimal interpolation–based quality control procedure of NCEP (Woolen 1991). According to the procedure, which is standard for dropsonde observations at NCEP, the complex quality control (Gandin 1990) was not applied to the targeted data. The humidity observations, however, were all rejected (by assigning a mandatory bad quality mark to them) due to recurring quality problems.

Since the FASTEX, NORPEX, and CALJET field experiments, major changes have been implemented in the NCEP analysis–forecast system (Derber et al. 1998). Therefore, care should be taken when results presented here and in earlier papers by Szunyogh et al. (1999a,b) are compared.

The changes in the medium-range forecast model (MRF) include new parameterization schemes for the land surface interactions, cumulus convection, gravity wave drag, radiation and clouds, and a new prognostic variable for the ozone mixing ratio. The changes in the SSI include the assimilation of new satellite data, changes to the radiative transfer calculations for the assimilation of satellite data, a three-dimensional analysis for the ozone mixing ratio, a background error term that is allowed to vary by latitude, an improved handling of supersaturation and negative moisture values, a nonlinear minimization algorithm and a nonlinear external iteration procedure. For our application, presumably the most important new feature of the analysis is the time interpolation of the background forecast that is now extended over the second half, that is, the 6–9-h period of the analysis time window. In order to understand the importance of this change, first we recall that the analysis is a (dominantly linear) combination of a short-term forecast (sometimes called background field or first

guess) and observations taken in a given analysis time window. An important step of the analysis procedure is the computation of the observation increment, the difference between the observation and the background field at the location and time of each observation. At NCEP the analysis time window is 6 h. Prior to 1200 UTC 15 June 1998 the first guess was a 6-h MRF forecast valid at the central time of the analysis time window. The observation increments were computed by interpolating the first guess in space, and also in time, but only for the first half of the analysis time window. For the second half of the time window the 6-h forecast was used without time interpolation. This approximation introduced negligible errors to the analysis in general since the vast majority of traditional weather observations are taken at the middle of the time window, thus no time interpolation is needed; and changes in the atmosphere at the synoptic scales are usually small for a 3-h period. Airborne reconnaissance observations, however, frequently span periods even longer than the analysis time window and target areas of rapid atmospheric developments. In this case, large phase errors may be introduced in the analysis by the assimilation of dropsonde data taken in the second half of the analysis time window as shown by Szunyogh et al. (1999a, Fig. 2) for the FASTEX dataset. Since 1200 UTC 15 June 1998, NCEP uses 3-, 6-, and 9-h short-term forecasts in its analyses scheme. The background field is now computed using linear interpolations of the 3- and 6-h forecasts for the first half of the analysis time window, and the 6- and 9-h forecasts for the second half.

The forecasts using targeted data tend to benefit from improvements in the analysis scheme. During former field experiments we observed a strong initial adjustment process (Szunyogh et al. 1999a): the surface pressure analysis increment (the difference between the analysis and the background) associated with targeted data started growing only after a 12–24-h period of initial decay. The same phenomenon was not observed during WSR99, which we attribute to the use of the new time interpolation scheme and changes in the background error term of the analysis.

### c. The control cycle

In order to evaluate the effects of the dropsonde data, a control analysis–forecast cycle was set up and run in parallel to the operational T62 analysis–forecast cycle. The control analysis cycle was identical to the operational T62 cycle except it used no dropsonde data. Since most of the dropsonde data were collected in time windows centered around 0000 UTC, the largest differences between the operational and control analyses were seen at this time. One daily medium-range forecast was run from the 0000 UTC control analysis, hereafter referred to as the control forecast.

The rejection of the dropsonde data was done after the quality control (QC) of the observed data. This pro-

cedure ensures that the only difference between the two datasets was the presence (lack) of dropsonde data in the operational (control) analysis.

The setup of the analysis–forecast experiments described here is identical to that followed for evaluating the effects of soundings from the NORPEX and the CALJET field experiments. For the evaluation of the FASTEX observations we used a slightly different approach. In FASTEX there were no parallel cycles. Instead, two analyses were created for each analysis time when dropsonde observations were taken: one with and another without the dropsonde data, both using the same operational background forecast. The main difference between the two approaches is that, when the analysis is cycled, dropsonde data assimilated at earlier times can also contribute to the differences in the latest analysis and ensuing forecasts through the differences in the background fields. This time we chose the cycling strategy because our goal was to explore the value of targeted observations in an operational environment. Note also that analysis and forecast data from the operational cycle were available at no extra computational cost.

It is important to note that the rejection of all dropsonde data from the control cycle does not mean that no observed data were assimilated into that cycle over the northeast Pacific. Along with reports from commercial aircraft and other in situ measurements a huge amount of satellite-derived information was also assimilated into both the control and the operational cycles. This satellite information includes 1) high-density wind vectors derived from infrared cloud-drift, water vapor, and cloud-top measurements of the *Geostationary Operational Environment Satellite-8* and *-10* (*GOES-8* and *-10*) by the National Environmental Satellite, Data, and Information Service (NESDIS); 2) low-resolution wind vectors derived from *GOES-8* picture triplet information by NESDIS, and from high- and low-level cloud-drift Geostationary Meteorological Satellite soundings from the Japan Meteorological Agency; 3) different radiances from the *GOES-8* satellite; and 4) radiance provided by the Microwave Sounding Unit of the polar-orbiting *NOAA-11* and *-14* satellites and the High Resolution Infrared Radiation Sensor of *NOAA-14*.

## 4. Flow regimes during WSR99

### a. Persistent structures in the circulation

The analysis of the flow regimes presented in this section is based on the daily 0000 UTC operational T126 analyses of NCEP between 13 January and 10 February. We use traditional analysis techniques explained in detail by James (1994; hereafter J94). Figures presented in J94 were also used as a basis for comparison between the WSR99 flow regimes and climatology. The time mean of an arbitrary quantity  $Q$  for the investigated time period is denoted by an overbar as  $\overline{Q}$  throughout this section.

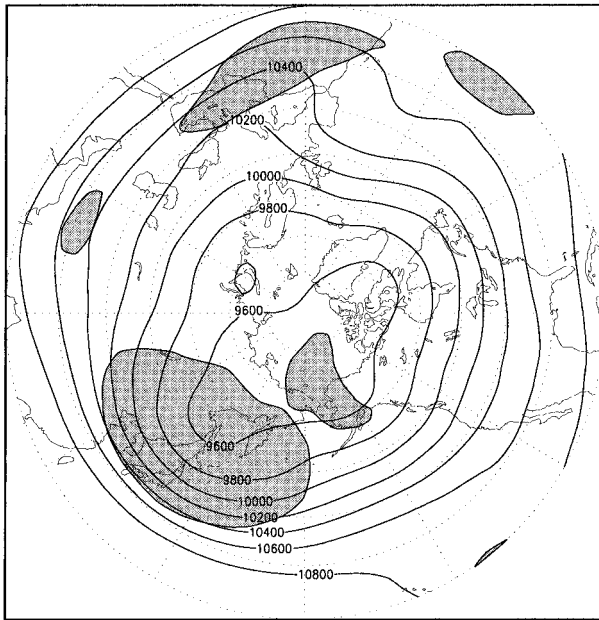


FIG. 1. The 0000 UTC time mean geopotential height of the 250-hPa surface for 13 Jan to 12 Feb 1999. Negative anomalies exceeding 100 m from the zonal mean are shaded.

Figure 1 presents the time mean geopotential height of the 250-hPa atmospheric level. This figure shows a strong zonal flow over the Pacific with a maximum gradient over and east of Japan. While the main features resemble climatology (cf. Fig. 6.1 of J94), there are some interesting deviations from it. Most interestingly, the trough over Japan is well pronounced, but the typical trough over Canada is now shifted to the east. A smaller-amplitude trough is also present over Alaska. This feature is even more obvious if we subtract the zonal mean from the time mean values (shown in shades in Fig. 1.) A comparison of the anomalies from the zonal mean to their long-term average (cf. J94, Fig. 6.2.a) shows that during WSR99 a weak negative anomaly took the place of a well-pronounced positive anomaly over Alaska and the Bay of Alaska.

#### b. The transient behavior of the circulation

In what follows, we will investigate the relationship between the dynamics of the Pacific storm track and the space and time propagation of the effect of dropsondes. Keeping in mind that the dynamical model of storm tracks (Chang and Orlanski 1993) emphasizes the importance of downstream development, or in other words, the role interactions between lower-level baroclinicity and upper-level eddy propagation play in the baroclinic energy conversion, a series of diagnostics were prepared to analyze this process for the WSR99 period. Here, we compute time-filtered temperature fluxes to detect the main regions of baroclinic energy conversions: poleward temperature fluxes convert the available potential

energy of the zonal mean flow to available eddy potential energy, while upward temperature fluxes convert the available eddy potential energy to eddy kinetic energy. The Eady index is used to measure the lower-troposphere baroclinicity, while the eddy kinetic energy of the 250-hPa level and a Hovmöller diagram for the meridional component of the wind at the 300-hPa level is used to detect the eastward-propagating eddies along the jet. The Hovmöller diagram will also help us to analyze the individual targeting cases.

Eddy statistics of the Northern Hemisphere circulation were computed after applying a high-pass time filter to the analyzed meteorological fields. That is, first we computed the 7-day running mean

$$Q_n^F = \frac{1}{7} \sum_{i=-3}^3 Q_{n+i}, \quad (2)$$

where  $Q_n$  denotes an analyzed quantity at the  $n$ th day in the series of daily analyses. Then a time series of daily eddy quantities,  $Q'_n$ , were generated by taking the deviation of the daily values from the corresponding running mean as

$$Q'_n = Q_n - Q_n^F. \quad (3)$$

Daily eddy quantities were used to analyze the eddy fluxes related to the baroclinic life cycles of synoptic-scale storms. Filtered transient eddy statistics are shown in Fig. 2 for the time period considered in this study. The zonal and meridional wind components, the temperature and the vertical velocity in pressure coordinate system are denoted by  $u$ ,  $v$ ,  $T$ , and  $\omega$ , respectively. The overlapping maxima in the northward (Fig. 2a) and upward (Fig. 2b) temperature fluxes at the 700-hPa pressure level around latitude  $40^\circ\text{N}$ , and the associated positive gradient in the transient eddy kinetic energy (Fig. 2c), suggest that the main regions of available potential to eddy kinetic energy conversion are over the North Pacific between  $150^\circ\text{E}$  and  $150^\circ\text{W}$ , from the midwest to the eastern United States and Canada, and the Atlantic Ocean just to the east of the North American continent. These regions, except for the Midwest, are overlapping with areas of the strongest low-level baroclinicity shown by Fig. 2d, which presents the average 24-h amplification

$$\lambda = e^\sigma \quad (4)$$

of the most unstable baroclinic mode for the atmospheric layer between 700 and 850 hPa. The computation of the mean 24-h growth rate,  $\sigma$ , following J94, is based on the Eady model; that is,

$$\sigma = 86\,400 \times 0.31f \overline{\frac{1}{N} \frac{\partial u}{\partial z}}, \quad (5)$$

where  $f$  is the Coriolis parameter,  $N$  is the Brunt–Väisälä frequency,  $u$  is the zonal component of the wind, and the vertical partial derivative is approximated by finite differences between the 700- and the 850-hPa levels.

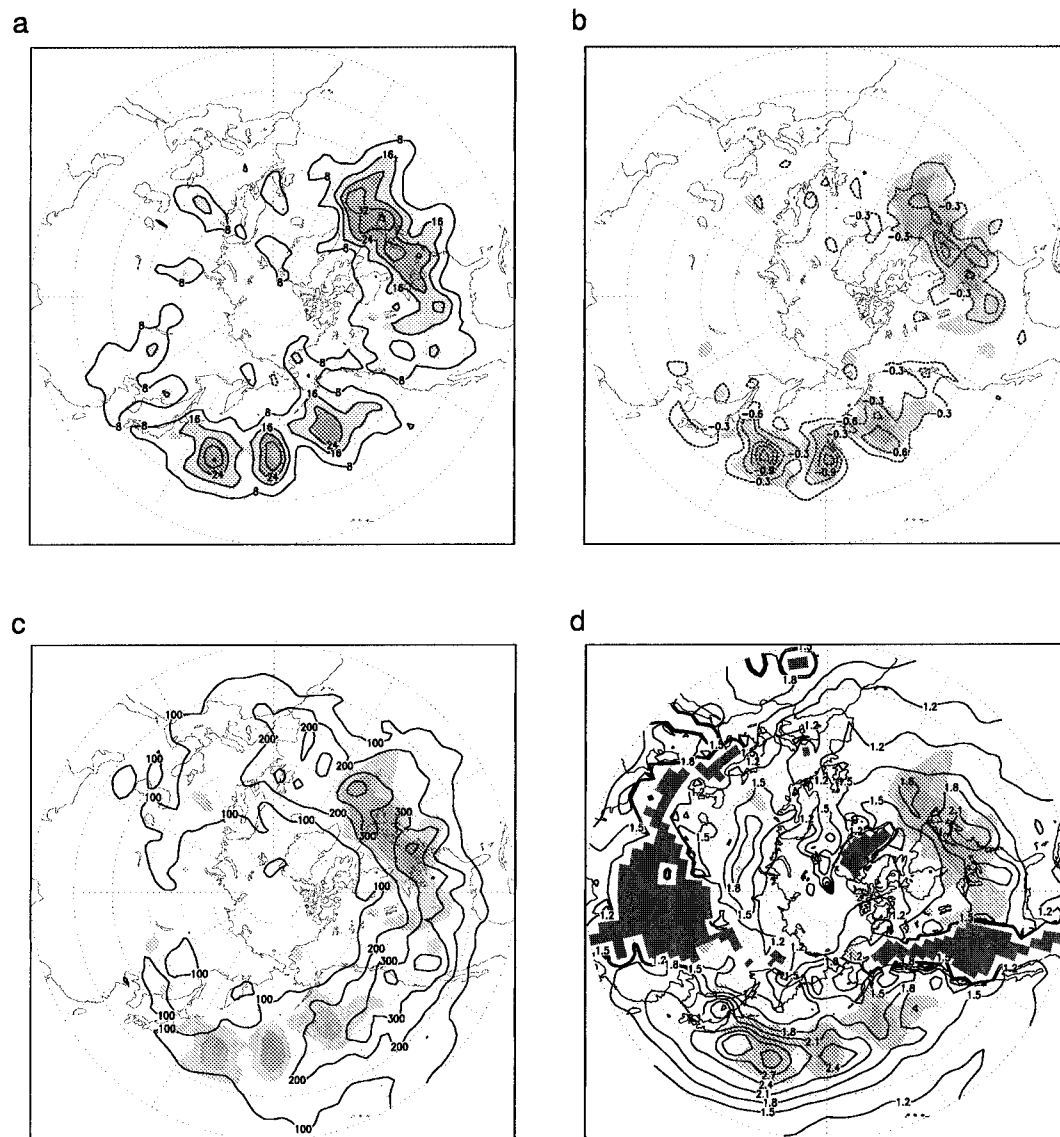


FIG. 2. High-pass-filtered eddy statistics. (a) Meridional temperature flux,  $\overline{v'T'}$  at the 700-hPa pressure level. Contour interval is  $8 \text{ K m s}^{-1}$ . (b) Vertical temperature flux  $\overline{w'T'}$  at the 700-hPa pressure level. Contour interval is  $0.3 \text{ KPa s}^{-1}$ . (c) Eddy kinetic energy,  $(u'^2 + v'^2)/2$  at the 250-hPa level. Contour interval is  $100 \text{ m}^2 \text{ s}^{-2}$ . (d) The 24-h amplification factor for the most unstable Eady mode at the 775-hPa pressure level. Mountainous regions are covered by black. Gray shades in all panels represent regions where the poleward eddy temperature flux is larger than  $16 \text{ km s}^{-1}$  (light gray) and  $24 \text{ km s}^{-1}$  (dark gray).

The local maxima of the eddy kinetic energy over the Rockies and on the western side of the Cascades, where there is a meridionally elongated maximum in the upward temperature flux as well, are presumably associated with orographic forcing. Notice also that the general geographical distribution, but not the details, of the eddy statistics is in good agreement with their climatology (cf. J94, Fig. 7.8).

Figure 3 presents the time–latitude diagram for the zonal component of the wind at the 250-hPa pressure level. The zonal average is shown for the  $140^{\circ}\text{W}$ – $180^{\circ}$  longitude band, where the vast majority of the sondes

were deployed. The position where each sonde splashed into the ocean surface is marked by a cross. Figure 4 is a Hovmöller (time–longitude) diagram that shows the meridional average of the meridional component of the wind at the 300-hPa pressure level for the  $30^{\circ}$ – $60^{\circ}\text{N}$  latitude band. The 300-hPa level and the meridional component of the wind was chosen based on an earlier study (Chang 1993) that demonstrated that eastward-propagating eddies associated with baroclinic downstream development are the easiest to detect by making these choices. The crosses (circles and squares) show the dropsonde locations (the center of the verification



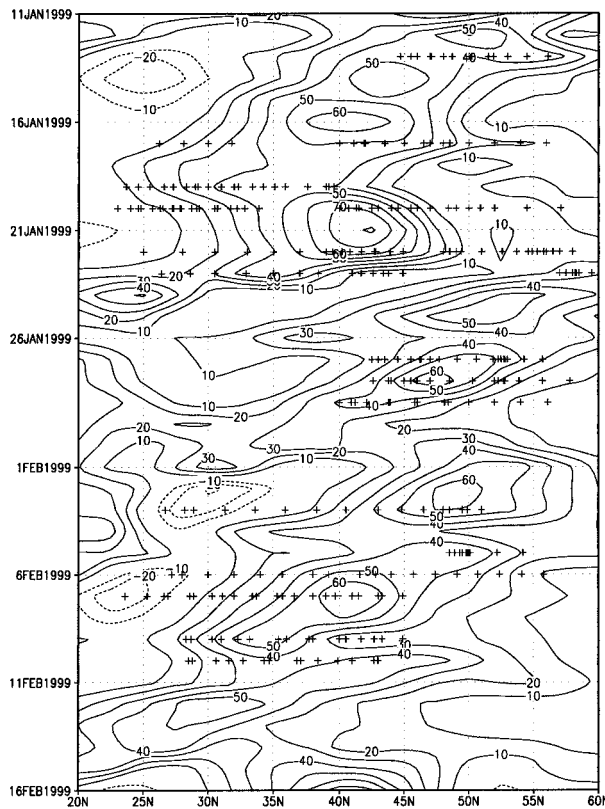


FIG. 3. Time-latitude cross section for the zonal mean of the zonal wind component between  $140^{\circ}\text{W}$  and  $180^{\circ}$  at the 250-hPa pressure level. Contour interval is  $10\text{ m s}^{-1}$ . The dropsonde locations are shown by crosses.

regions), while straight lines indicate the connection between the dropsonde missions and the verification regions. A time-latitude diagram was also prepared (but not shown) for the zonal average of the meridional wind component in order to detect changes in the meridional wind component of the jet.

The time-latitude and the Hovmöller diagrams suggest the existence of three well-distinguishable regimes: 1) eastward-propagating waves along a strengthening southwesterly, later westerly, jet centered around  $43^{\circ}\text{N}$  during 13–21 January; 2) a dominantly nonzonal regime south of  $45^{\circ}\text{N}$  and eastward-propagating waves along a weaker jet centered around  $50^{\circ}\text{N}$  between 22 January and 3 February; and 3) a northwesterly jet shifting from  $45^{\circ}\text{N}$  to  $30^{\circ}\text{N}$  during 4–13 February. The Hovmöller diagram shows well-distinguishable downstream development on eight occasions (marked by shaded areas). The eddy pockets propagate, on average, at a rate of  $30^{\circ}\text{ day}^{-1}$  in contrast to the  $10^{\circ}\text{ day}^{-1}$  average speed of the individual systems. These speeds are in a good agreement with those usually observed for the Northern Hemisphere winter (A. Persson 1999, personal communication).

For the East Coast verification regions the ensemble-based targeting techniques anticipated a close to  $30^{\circ}$

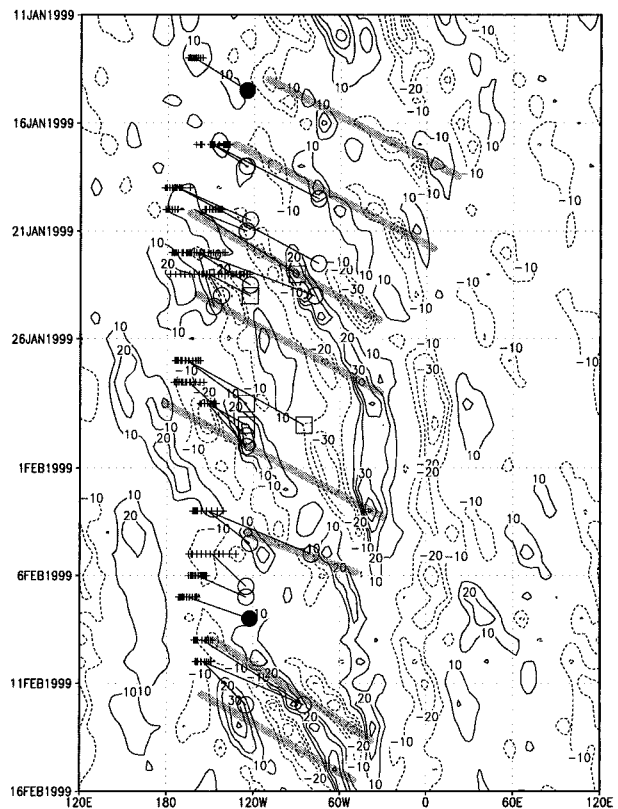


FIG. 4. Hovmöller diagram (time-longitude cross section) for the meridional mean of the meridional wind component between  $30^{\circ}$  and  $60^{\circ}\text{N}$  at the 300-hPa pressure level. Contour interval is  $10\text{ m s}^{-1}$ . The dropsonde locations are shown by crosses, while the eastward-propagating wave packets are marked by shaded areas. Open circles (closed circles, open squares) identify the center of the verification regions for cases where the quality of the forecast within the verification region was improved (was not changed, was degraded) by the use of targeted data.

$\text{day}^{-1}$  propagation speed. All missions targeting East Coast forecasts, except the flight on 27 January, were associated with downstream developing wave packets. On 27 January, dropsonde data were collected and a downstream impact was expected in a gap between two packets of eastward-propagating waves. One should keep in mind, however, that the selection of the verification region and the sensitivity analysis-flight planning is always based on a 2-day forecast. Hovmöller diagrams (not shown) based on forecasts from 25 and 27 January revealed that a wave packet was predicted to connect the targeting region on 27 January and the verification regions 48 and 72 h later. The forecast was changed at the intended location, since the wave packet was still forecast on 27 January, but as it will be shown later, the effect of the targeted data on forecast quality was seriously compromised. Note also that no East Coast verification regions were selected for the three missions between 5 and 7 February, a period for which no downstream development can be detected.

The picture is far more colorful for the West Coast

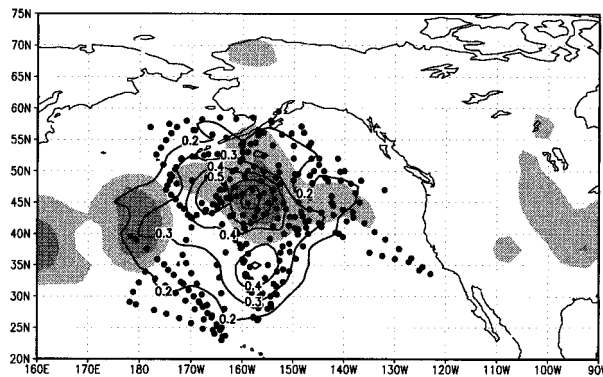


FIG. 5. Composite mean surface pressure signal (hPa) at analysis time for the 15 flight days (contour lines). Shades are as in Fig. 2. Full circles indicate dropsonde locations.

and Alaskan verification regions. There are still cases, for which the sensitivity analysis selected locations along the wave packets (the best example is 20 Jan), but on several occasions the land-falling system itself was observed over the ocean (examples are the systems with Alaskan verification regions, 13 Jan, 17 Jan, and 3 Feb).

## 5. The analysis–forecast effect of targeted data

### a. Signal from the targeted data

In order to describe the propagation of the effect of dropsonde data in space and time a new set of Eulerian variables is defined by the absolute value of the difference between the prognostic variables in the operational and the control forecasts. These new variables are distinguished from variables in the forecasts by using the term *signal*.

### b. Analysis effect of the targeted data

Figure 5 presents the composite mean of the surface pressure signal at analysis time for the 15 flight days. The largest mean analysis signal occurs in a region around 45°N, 160°W, where a large number of dropsondes sampled the easternmost local maximum in the northward (and upward) temperature fluxes over the Pacific (Figs. 2a,b). It is also interesting to note that in the main regions of baroclinic energy conversion (large temperature fluxes) even a few dropsondes could make substantial changes. As to the analysis effect of the dropsonde data in the individual cases, there is a group of “big impact” cases (13 Jan, 5 Feb, and 10 Feb) for which the maximum change in the surface pressure exceeds 2.0 hPa, a number of “small impact” cases with changes in the analysis less than 1 hPa (19, 27, and 28 Jan and 5 Feb), while for the rest of the cases the maximum impact is between 1.0 and 2.0 hPa. We also looked at changes in the analyzed height of the 500-hPa level. Larger (smaller) changes tend to happen in cases, for

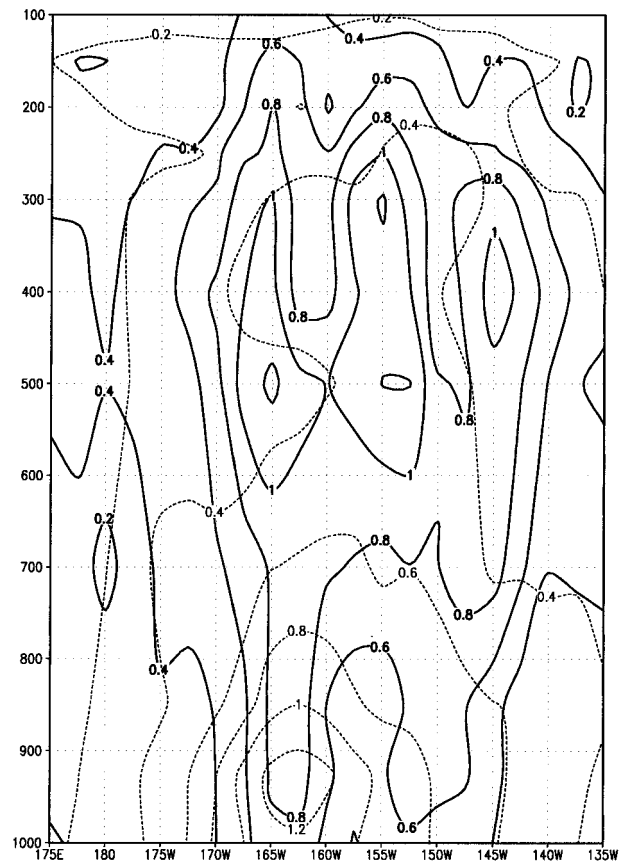


FIG. 6. Height–longitude cross section for the composite mean of the meridional wind component signal (solid, contour interval is 0.2  $\text{m s}^{-1}$ ) and the virtual temperature signal (dashed, contour interval is 0.2 K) for the 15 flight days at 45°N latitude.

which changes in the surface pressure were also larger (smaller), although there are exceptions from this rule. The big impact cases (maximum change is larger than 20 m) are 17 January, 5 February, and 10 February, while the small impact (maximum change is smaller than 10 m) cases are 19 and 27 January.

In Fig. 6 a height–longitude cross section of the meridional wind and temperature signal is shown at 45°N, where the largest surface pressure signal and the most intense baroclinic energy conversion were observed. The largest changes in the wind component are centered around 155°W (the position of the maximum in the surface pressure signal and of the most intense baroclinic energy conversion) and they occur in the upper half of the troposphere with maxima between the 600- and 250-hPa pressure levels. The largest changes in the temperature are between 160° and 165°W, and they are concentrated at the lowest layers of the troposphere, with a secondary maximum in the layer between 300 and 400 hPa.

The positive effect of the dropsondes on the analyses can be inferred from the fact that they improved the first guess forecasts at later analysis times (Fig. 7a).

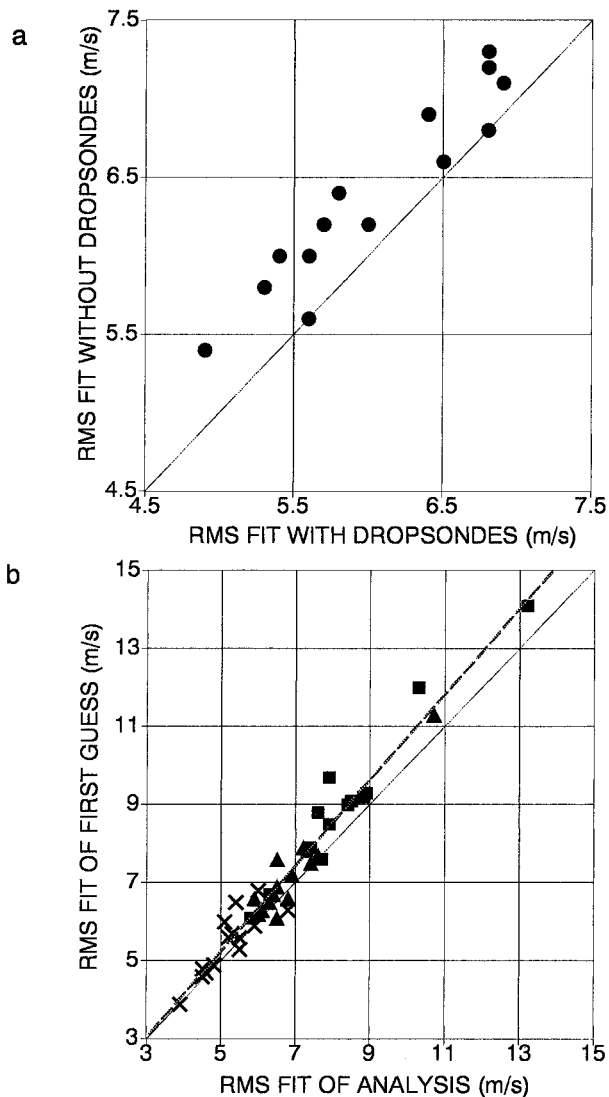


FIG. 7. (a) Rms fit of 6-h first-guess wind forecasts (accumulated for the 1000–250-hPa layer) from the operational (horizontal axis) and the control analysis–forecast cycles for the last 14 WSR99 flights (the first-guess forecasts were identical for the first case and therefore are not shown). (b) Fit of the control first guess (vertical axis) and the analysis (horizontal axis) to independent dropsonde wind observations. The crosses, triangles, and squares show results averaged for the 1000–700-, 700–400-, and 400–250-hPa layers, respectively.

Figure 7a shows the fit of the 6-h forecasts, which were valid at the times when dropsonde data were assimilated to the operational cycle, to the dropsonde observations. We note that the dropsonde observations are independent verification data for both cycles in this case, since they were assimilated into the operational cycle at a time later than when the 6-h forecasts were initiated. In 12 of the 14 cases the operational 6-h forecast was superior to the control due to the positive influence of *past* dropsonde data. This improvement is statistically highly significant (at the 0.05% level, see the appendix for details), and amounts to an average 6% improvement

in the rms fit of the 6-h forecasts to the wind observations.

Figure 7b demonstrates that the analysis and the 6-h forecast errors in the control cycle are highly correlated ( $r = 0.98$ , statistically significant at a lower than 0.05% level), indicating that the primary sources of analysis errors over the North Pacific are errors in the background forecasts. The derivative of the fitted line is slightly larger than 1, which means that, as expected, nontargeted (mainly satellite) data over the Pacific are somewhat more efficient in correcting large- than small-magnitude errors in the background forecasts.

It is important to note that the fit of the operational analyses (using all dropsonde data) was worse in cases when there were large errors in the 6-h operational forecasts used as first guesses for these analyses (not shown). This phenomenon can be explained by the fact that the analysis scheme assumes static (time independent) background error variances, and in the presence of unusually large first-guess errors it cannot introduce sufficiently large corrections in the direction of the observed data. The latest changes in the SSI will make it possible for the background error variance to vary with the synoptic situation in the future (Derber et al. 1998). We anticipate that an adaptive background error term, once implemented, will further increase the value of targeted observations in improving numerical weather predictions.

### c. Time evolution of the mean signal

The time evolution of the mean surface pressure signal for the 15 flight days is shown in Fig. 8. At the beginning, the dominant signal moves toward the east by deepening along the core of the storm track. Note that in 13 out of 14 West Coast targeting cases the center of the verification area was at 122.5° or 125°W, and the average position of the center for all West Coast verification regions was 43.2°N, 124.1°W. In Fig. 8 this average position of the West Coast verification area is shown at 36-h lead time, which is the average verification lead time for the West Coast cases (Table 1). This means that the maximum of the mean signal reached the mean West Coast verification area at the average lead time. A similar conclusion can be drawn for the individual cases: the maximum signal in the surface pressure and the precipitation forecasts is within the preselected verification region at the preselected verification time (24–60 h) for all 14 individual cases.

For longer than 36-h lead times the signal propagates farther to the east, first over the North American continent and then over the Atlantic. In most cases, the maximum signal was within the Alaska and East Coast verification regions, but drawing a composite map for these cases would be meaningless due to the large diversity in verification times and locations. The eastward propagation speed of the signal is about 30° day<sup>-1</sup>: the position of the maximum in the leading part of the signal is 160°W (130° and 100°W) at 12-h (36 and 60 h) fore-

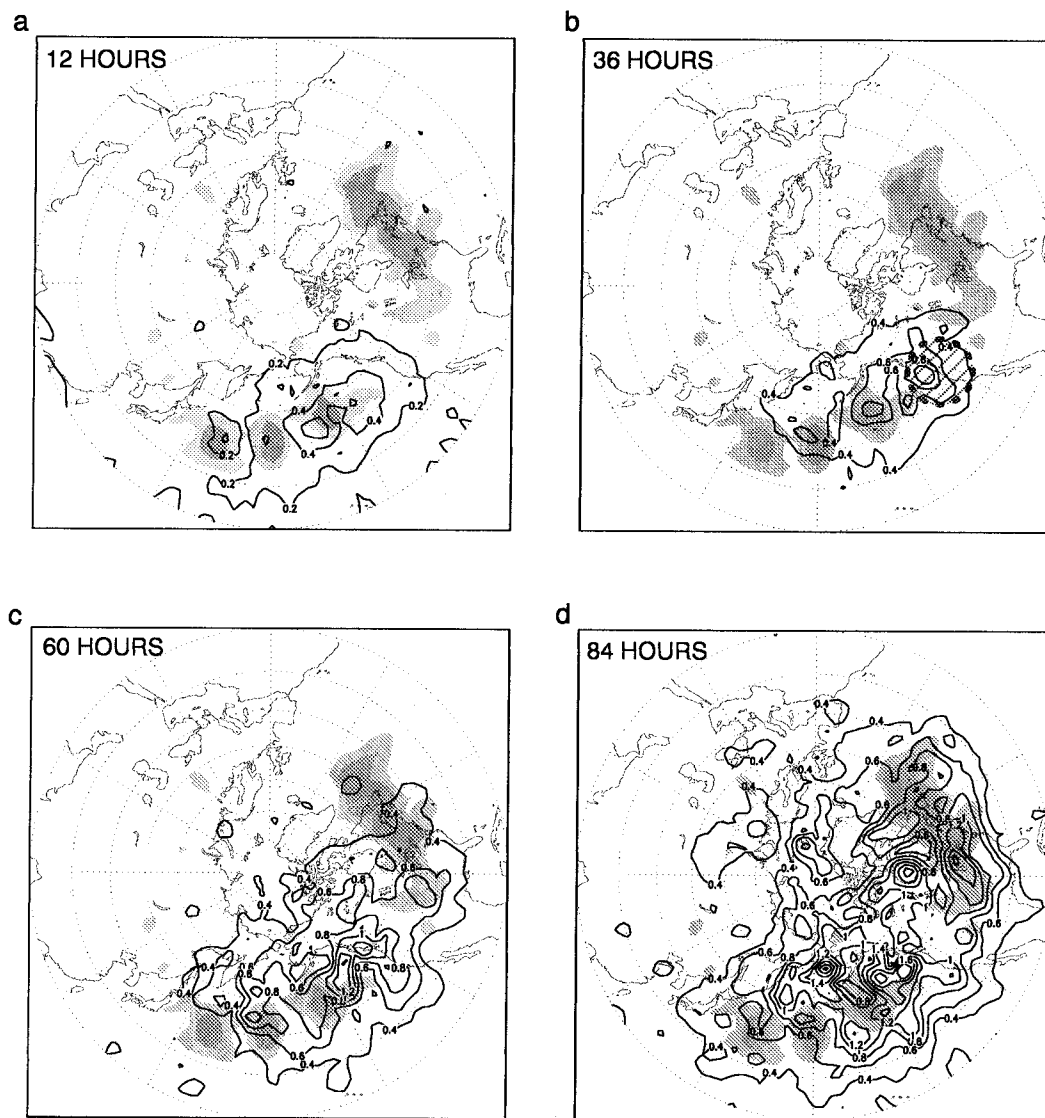


FIG. 8. Composite mean of the surface pressure signal for the 15 flight days at (a) 12-, (b) 36-, (c) 60-, and 84-h (d) forecast lead times. Contour interval is 0.2 hPa at 12, and 0.4 hPa at longer lead times. Contour shades are as in Fig. 2.

cast lead times, while the leading edge of the signal is at  $120^{\circ}\text{W}$  ( $60^{\circ}\text{W}$ ) at 12-h (60-h) forecast lead time. In addition to the propagation speed, the general behavior of the signal is also in a good agreement with the general model of storm tracks based on the conceptual model of baroclinic downstream development. In regions where the low-level baroclinicity is weak, the surface signature of the signal is also weak. In contrast, it deepens rapidly in areas of stronger baroclinicity and more intense energy conversion processes.

At lead times of 60 h or more after the targeted observations were deployed, several new local maxima also appear on the northern side of the storm track. Beyond 84 h, these northern patterns dominate the signal. The position of these maxima have no obvious relationship to the main baroclinic regions. Instead, they

tend to appear in regions of negative zonal anomalies of the time mean flow indicating a closer relationship to interactions between the high-frequency transients and the low-frequency variability of the atmosphere.

## 6. Verification of forecasts

### a. Time evolution of forecast errors

The time evolution of the *estimated mean error*,

$$\frac{1}{15} \sum_{i=1}^{15} |f_i^c - a_i^c|, \quad (6)$$

in the control forecast for the 15 flight days is shown in Fig. 9. Here, the estimates of the daily forecast errors are computed as the absolute value of the surface pres-

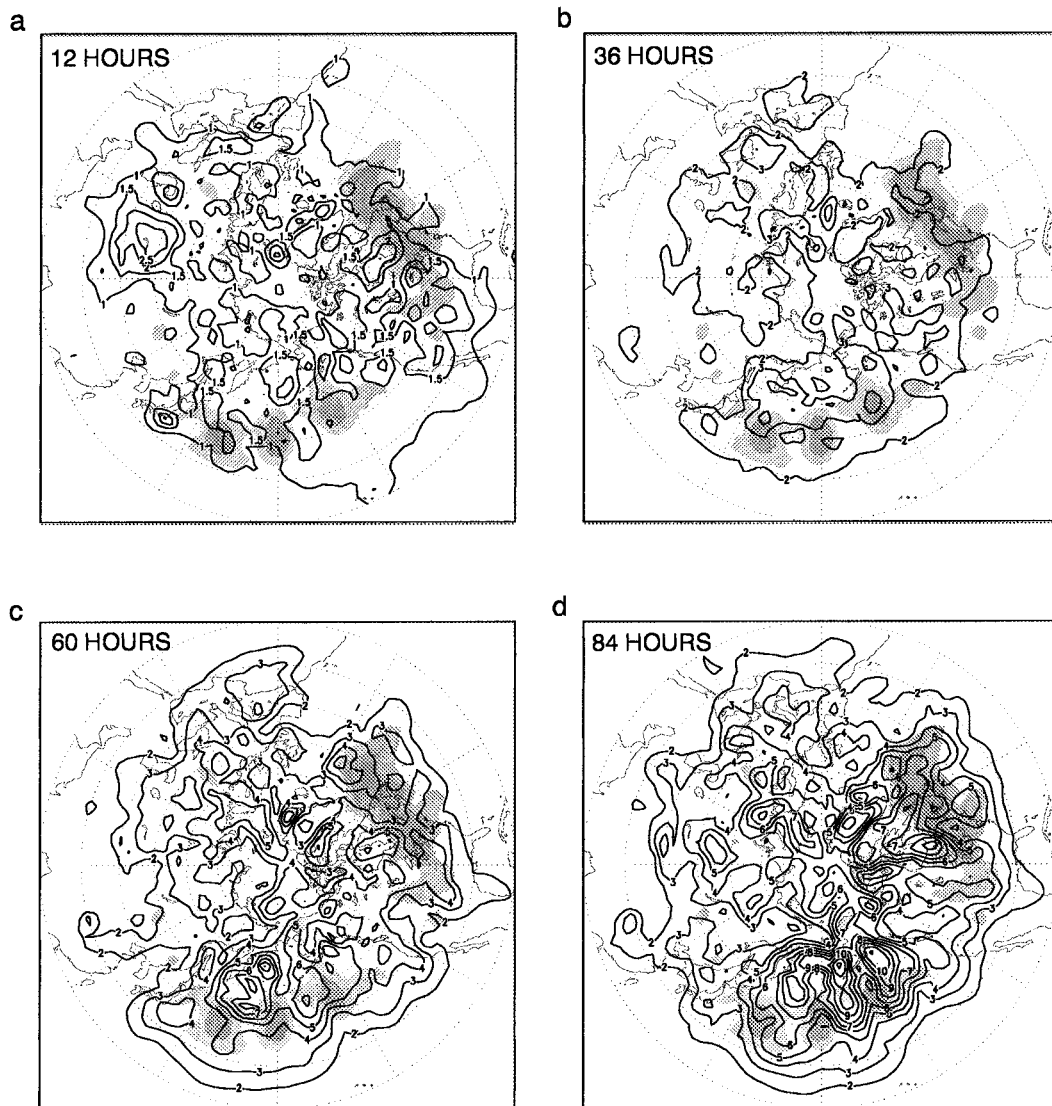


FIG. 9. Same as Fig. 8 except for the analysis-based estimate of the forecast error. Contouring starts at 1 hPa with an interval of 0.5 hPa for the 12-h lead time, and at 2 hPa with 1-hPa intervals for longer lead times.

sure difference between the control forecasts ( $f_i^c$ ) and the control analyses ( $a_i^c$ ). The limitation of this otherwise standard verification technique is that the analyses used for verification contain errors; the short-term verification results in particular should be handled with care.

The largest short-term (12–24-h) forecast errors occurred in the main baroclinic regions and in orographic areas. In the 36–60-h forecast lead time range errors are organized around the baroclinic regions, while for longer lead times the dominant errors are shifted to north of the storm track.

We can conclude that the difference between the mean amplitudes of the signal and the error is not overwhelming. The mean forecast error in the average West Coast verification region at 36-h lead times is 3.3 hPa, only three times larger than the maximum of the mean signal.

This means that the signal on average was large enough to cause substantial changes in forecast performance.

#### b. Verification against analyses

We define the analysis-based estimate of the *forecast error reduction* as the difference between the analysis-based estimates of the error in the control and operational forecasts ( $f_i^o$ ):

$$\frac{1}{15} \sum_{i=1}^{15} (|f_i^c - a_i^c| - |f_i^o - a_i^o|). \quad (7)$$

Wherever this quantity is positive (negative) the forecast is improved (degraded) by the targeted data. The time evolution of the average forecast error reduction is shown in Fig. 10. For short (12 h) lead times the re-

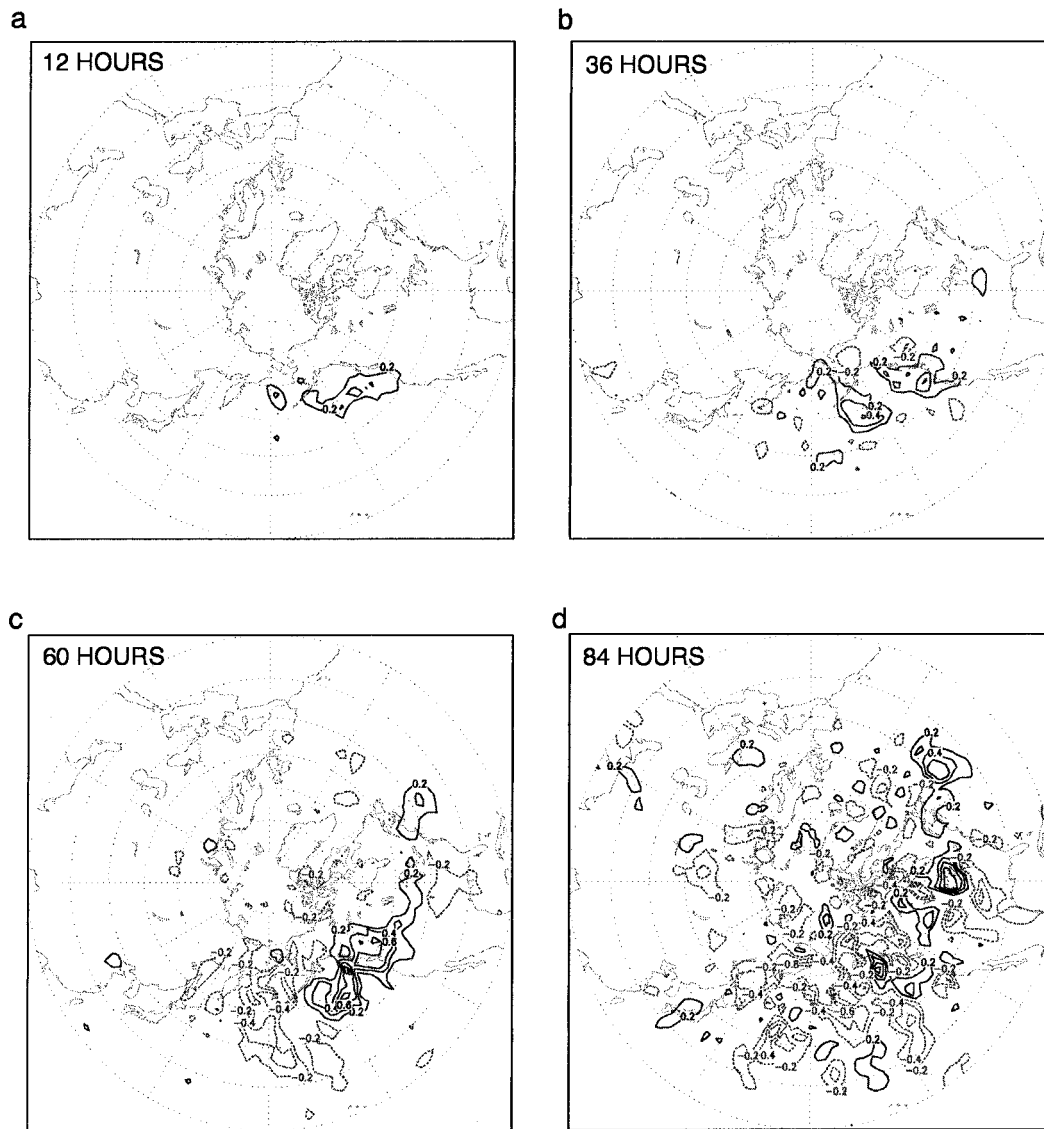


FIG. 10. Same as Fig. 8 except for the analysis-based estimate of the forecast error reduction. Contour interval is 0.2 hPa.

duction in the average error seems to be unquestionable, though a careful comparison of Figs. 8 and 10 at 12-h lead time shows that in regions where the signal has its maximum there is no change in forecast quality. This can happen if added observations change the forecast without changing their quality or if there is a difference between the two verifying analyses similar to that between the forecasts since

$$\begin{aligned} & \frac{1}{15} \sum_{i=1}^{15} \left| |f_i^c - a_i^c| - |f_i^o - a_i^o| \right| \\ & \leq \left| \frac{1}{15} \sum_{i=1}^{15} (f_i^c - f_i^o) - \frac{1}{15} \sum_{i=1}^{15} (a_i^c - a_i^o) \right| \approx 0. \quad (8) \end{aligned}$$

A comparison of the verifying operational and control

analyses revealed that the forecast and the analysis differences were similar within the region where the signal had a maximum.

For lead times longer than 24 h there is a good consistency between the regions of large changes in the forecast quality and the positions of the maximum signal. A comparison between Figs. 9 and 10 shows that the region where the error reduction exceeds 10% propagates to the east along the storm track as the verification time is increasing. Interestingly, a contiguous area of degradation appears at 60–84 h lead time north and south of, and especially behind the leading edge of, the signal. Beyond 84 h (not shown), except for the leading edge of the signal, where the improvement is consistent up to 5 days, the apparent boundary between regions of improvement and degradation gradually disappears.

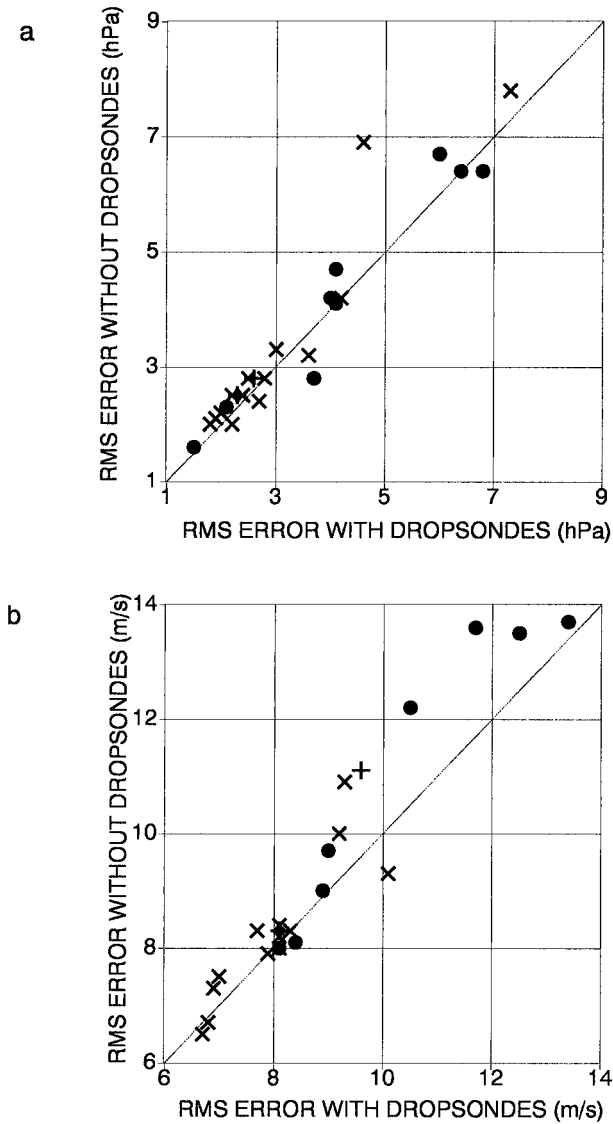


FIG. 11. Rms error (measured against observations) in (a) the surface pressure and (b) winds forecasts for the operational (horizontal axis) and the control (vertical axis) forecasts in the preselected West Coast (dots), East Coast (crosses), and Alaskan (plus signs) verification regions.

*c. Verification against data*

Individual forecasts for the dates and time periods shown in Table 1 have been verified against radiosonde observations from the verification regions. Forecast rms error scatterplots are shown in Fig. 11 for the surface pressure and the wind speed between the 1000- and the 250-hPa atmospheric levels. Crosses (dots, squares) above the 45° line indicate improvement in the West Coast (East Coast, Alaskan) verification regions, while symbols below the same line show forecast degradation. The surface pressure forecasts were improved/degraded in 9/3 (5/2) out of the total number of 14 (9) cases for the West Coast (East Coast). For wind forecasts, the

number of cases with improvement/degradation are eight/four (six/three) for the two regions, respectively. On the two occasions when the verification region was over Alaska both the surface pressure and the wind forecasts were improved by the extra data.

Scores for the individual forecasts are presented in Table 1. The precipitation scores are based on a subjective comparison between the predicted and the analyzed 12-h precipitation amounts, where the latter is based on rain gauge observations over the continental United States. The main evaluation criteria were whether extra data could improve 1) the timing of the precipitation event and 2) the predicted precipitation amounts. The impact was judged to be positive (negative) if the forecast was either improved (degraded) with respect to both criteria or if it was improved with respect to one criterion without a substantial impact on the other. The summary scores were computed in a similar way: +1 (−1 and 0) was assigned to the forecast in each forecast verification category for which it improved (degraded, had neutral impact on) the prediction. Then the values were added up for each forecast day and the summary score was judged to be positive (negative or neutral) if the sum was positive (negative or zero). It is convincing that the different forecast scores were improved in most cases and the overall forecast quality was enhanced in 18 out of the total number of 25 cases. The forecast improvement for the summary (surface pressure, wind, and precipitation) score is statistically significant at the 0.5% (1%, 10%, 0.5%) level.

A comparison of Table 1, Fig. 3, and Fig. 4 shows that four out of the five degraded forecasts are associated with three reconnaissance flights during times of nonzonal flow or regime transitions south of 45°N over the northeast Pacific. There was only one case (20 Jan 1999) when sampling the zonal regime degraded a forecast. The difference in the success rate of targeting during zonal and nonzonal regimes is even more striking if we recall that more missions took place during when zonal (nine) than nonzonal (six) regimes were dominating the flow south of 45°N. We speculate that this may be due to the fact that the 2-day forecasts used in the sensitivity analysis have considerably more uncertainty for the nonzonal regime. As was already pointed out in section 4b the flight on 27, February, which produced degradations for both the West and East Coasts, was intended to target a wave packet that has never existed in reality. In these cases the dynamical influence of the targeted region can be much weaker on the flow in the verification region at the verification time than as expected based on 2-day forecasts available at the flight planning time. In cases like these the very weak data impact on forecast quality can either be positive or negative by mere chance. We also add that we are not aware of any plausible reason why the assumptions in the SSI analysis scheme would be less accurate for nonzonal flows.

## 7. Conclusions

In this paper we evaluated the ability of the operational NCEP MRF analysis–forecast system to extract useful information from the targeted dropsonde data collected during the quasi-operational WSR99 program. In order to gain a better understanding of the underlying physical processes we attempted to explore the relationship between the dynamics of the storm track and the impact of dropsonde data. In summary we can draw the following conclusions.

- Large analysis errors over the northeast Pacific are associated with large first-guess errors. The dropsonde data tend to reduce these analysis errors, which are closely linked with areas of baroclinic energy conversion.
- The dominant signal from the dropsonde data propagates to the east along the storm track at an average speed of  $30^{\circ} \text{ day}^{-1}$ . This propagation speed along with the analysis of a Hovmöller diagram based on the 300-hPa meridional wind component suggests that downstream development plays an important role in spreading out the effect of dropsonde data from the targeted area.

Our results suggest that the 15 quasi-operational reconnaissance missions were successful in the following sense.

- The targeted data produced reasonably large surface pressure and precipitation signals in every case, typically with the maximum within the preselected verification region at verification time.
- In most cases the dropsonde data reduced the errors in the forecasts. Of those cases with a nonneutral impact on the forecast quality, nearly 80% of the forecasts were improved.
- The largest improvements occur where they are most needed, reducing, on average, by 10%–20% the forecast errors along the storm track. This result is based on the comparison of Figs. 9b and 10b. [See also Fig. 5 in Toth et al. (2000).]

While the overall forecast performance of our targeting strategy is encouraging, some of the results presented in this paper call for further investigation. In our small sample (15 cases), an intriguing positive correlation between forecast improvements due to targeted observations and the zonality of the flow was found. This result, if it can be confirmed by future research, may have important practical consequences. First, targeting may have more substantial forecast benefits in the prediction of a chain of baroclinic events developing along a strong zonal flow than predicted by earlier studies based on global forecast scores. Second, improvements in the initial conditions unrelated to the use of targeted data may also have a more positive impact on the prediction of severe winter storms than what traditional global verification scores may suggest. When

the forecast impact of analysis changes are tested, specific scores measuring the forecast performance along the storm tracks should also be considered, along with traditional global scores. For instance, we believe that the time interpolation of the background term in the SSI may have an important contribution to the success of our targeting efforts, even though earlier research conducted at the operational prediction centers found that the effects of such a change on global forecast scores are negligible.

Although the ETKF algorithm has proved to be a reliable tool in selecting dropsonde locations that can produce significant changes in the forecasts for the verification region, the technique can be further improved to produce quantitatively more accurate predictions for the forecast effect of adaptive observations. For example, the optimal choice of the verification variables for the ETKF computations is still an open problem. Likewise the optimal spacing between the dropsondes should also be explored. The ETKF may help to investigate the problem, but the use of the adjoint of the operational data assimilation system (Baker and Daley 1999) may also be required.

Based on the positive results collected during the NORPEX-98, CALJET, and WSR99 field programs we recommend that winter storm reconnaissance missions over the northeast Pacific be implemented into NWP operations on a regular annual basis. Our results suggest that a fully automated version of the ETKF technique would be a reliable practical tool to provide the information necessary for the identification of dropsonde locations that can produce on average 10%–20% forecast error reduction in the verification regions at 2-day forecast lead time. Results of data impact studies carried out with the operational analysis–forecast system of ECMWF suggest that a Pacific winter reconnaissance program would also have a positive forecast effect over Europe in the medium range and the Northern Hemisphere in general (Cardinali 1999; F. Boutier 1999, personal communication). A preliminary evaluation of forecasts from the WSR2000 field program showed, in terms of forecast improvements, results similar to those presented here and in the summary paper of Toth et al. (2000).

*Acknowledgments.* The 1999 Winter Storm Reconnaissance Program, and hence this study, would not have been possible without the work of a large number of participants and collaborators. First, we would like to acknowledge the dedicated work of the NOAA G-IV (led by Sean White) and the USAF Reserve C-130 flight crews (coordinated by Jon Talbot). Coordination with the flight facilities was provided by CARCAH, led by John Pavone. The WSR program benefited from a collaboration with the concurrently run Severe Clear-Air Turbulence Colliding with Air Traffic research program under the leadership of Cecile Girz (NOAA/ERL/FSL) and Mel Shapiro (NCAR). The European Centre for



Medium-Range Weather Forecasts is credited for providing their ensemble forecast data to be used in the sensitivity calculations in real time. The forecast cases were selected in real time by NWS field office and NCEP/HPC forecasters coordinated by David Reynolds. Mark Iredell, Jack Woollen, and Timothy Marchok provided valuable help with setting up the parallel analysis–forecast cycle, manipulating data, and creating graphics. The support and advice of Stephen Lord of EMC in organizing and running the field program was invaluable. Glen White, Milija Zupanski, Ralph Petersen, and one of the anonymous reviewers provided helpful comments on an earlier version of this manuscript. We are especially grateful to Anders Persson of ECMWF for his advice in the preparation of Fig. 4 and for sharing with us his thought-provoking notes on downstream development and potential vorticity thinking. The work of S. J. Majumdar and C. H. Bishop (B. J. Etherton and C. H. Bishop) was partly supported by the Grant NSF ATM-96-12502 (NSF ATM-98-14376).

## APPENDIX

### On the Reliability of the Forecast Verification Results

The estimated forecast rms error can be decomposed into three parts:

$$\overline{(f - o)^2} = \overline{(f - t)^2} + \overline{(o - t)^2} - 2\overline{(f - t)(o - t)}.$$

Here  $f$  ( $o$ , and  $t$ ) stands for the forecasts (observations and truth), the differences are taken at the observational locations, and the overbar denotes averaging over the number of observations. The first term on the right-hand side is the real rms error in the forecast, the quantity that we try to estimate. The second term, the rms error in the observations, and the third term, the covariance between the observational and the forecast errors, represent the error in the forecast error estimate. We note that the last term goes to zero as the number of observations involved in the verification is increasing, since none of the instrumental and the representativeness components of errors in the traditional (and not remotely sensed) observations are thought to be correlated with the forecast errors. Taking the difference between two forecasts verified against the same sufficiently large set of data we get

$$\overline{(f_1 - o)^2} - \overline{(f_2 - o)^2} = \overline{(f_1 - t)^2} - \overline{(f_2 - t)^2},$$

where the subscripts are to distinguish between the two forecast. The most notorious error component, the rms error in the observations, is canceled out completely by taking the difference. Hence with a sufficiently large number of observations forecast improvements much smaller than observational errors can be safely detected. Care should be taken, however, because when forecasts are verified in a limited geographical region against a limited set of observations, small but nonnegligible dif-

ferences between the two covariance terms may contaminate the results for the small-difference cases. In order to circumvent this problem significance tests have to be performed. Following standard procedure of mathematical statistics (see, e.g., Rao 1973) one can test whether the number of investigated cases was large enough to detect changes in the forecast quality at an acceptable confidence level. The significance levels presented in this paper are defined as the probability of the event that the detected changes in forecast quality appeared by mere chance.

## REFERENCES

- Baker, N. L., and R. Daley, 1999: Observation and background adjoint sensitivity in the adaptive observation targeting problem. Preprints, *13th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 27–32.
- Bergot, T., G. Hello, and A. Joly, 1999: Adaptive observations: A feasibility study. *Mon. Wea. Rev.*, **127**, 743–765.
- Bishop, C. H., and Z. Toth, 1996: Using ensemble to identify observations likely to improve forecasts. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 72–74.
- , and —, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765.
- Buizza, R., and A. Montani, 1999: Targeting observations using singular vectors. *J. Atmos. Sci.*, **56**, 2965–2985.
- , T. Petroligias, T. N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Weidi, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1935–1960.
- Cardinali, C., 1999: An assessment of using dropsonde data in numerical weather prediction. ECMWF Tech. Memo. 291, 20 pp.
- Chang, E. K. M., 1993: Downstream development of baroclinic waves as inferred from regression analysis. *J. Atmos. Sci.*, **50**, 2038–2053.
- , and I. Orlanski, 1993: On the dynamics of storm tracks. *J. Atmos. Sci.*, **50**, 999–1015.
- Derber, J., and Coauthors, 1998: Changes to the 1998 NCEP Operational MRF model analysis–forecast system. NOAA/NWS Tech. Procedure Bull. 449, 16 pp. [Available from Office of Meteorology, National Weather Service, 1325 East–West Highway, Silver Spring, MD 20910.]
- Emanuel, K. A., E. N. Lorenz, and R. E. Morss, 1996: Adaptive observations. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 67–69.
- Fischer, C., A. Joly, and F. Lalauette, 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. Wea. Rev.*, **118**, 2754–2767.
- Gelaro, R., R. H. Langland, G. D. Rohaly, and T. E. Rossmund, 1999: An assessment of the singular-vector approach to target observations using the FASTEX dataset. *Quart. J. Roy. Meteor. Soc.*, **125**, 3299–3328.
- Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity. *Quart. J. Roy. Meteor. Soc.*, **124**, 1225–1242.
- James, N. I., 1994: *Introduction to Circulating Atmospheres*. Cambridge University Press, 422 pp.
- Joly, A., and Coauthors, 1997: The Fronts and Atlantic Storm-Track Experiment (FASTEX): Scientific objectives and experimental design. *Bull. Amer. Meteor. Soc.*, **78**, 1917–1940.
- , and Coauthors, 1999: Overview of the field phase of the Fronts

- and Atlantic Storm-Track Experiment (FASTEX) project. *Quart. J. Roy. Meteor. Soc.*, **125**, 3131–3270.
- Langland, R. H., and Coauthors, 1999: The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Amer. Meteor. Soc.*, **80**, 1363–1384.
- Lord, S. J., 1996: The impact on synoptic-scale forecasts over the United States of dropwindsonde observations taken in the northeast Pacific. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 70–71.
- Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulations with a small model. *J. Atmos. Sci.*, **55**, 633–653.
- Morss, R. E., 1999: Adaptive observations: Idealized sampling strategies for improving numerical weather prediction. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 225 pp. [Available from UMI Dissertation Services, Ann Arbor, MI.]
- , K. Emanuel, and C. Snyder, 2000: Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.*, in press.
- Orlanski, I., and J. P. Sheldon, 1993: A case of downstream baroclinic development over western North America. *Mon. Wea. Rev.*, **121**, 2929–2950.
- , and —, 1995: Stages in the energetics of baroclinic systems. *Tellus*, **47A**, 605–628.
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747–1763.
- , —, R. J. Purser, W.-S. Wu, and Z.-X. Pu, 1997: The NCEP global analysis system: Recent improvements and future plans. *J. Meteor. Soc. Japan*, **75** (1B), 359–365.
- Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in chaotic systems. *Tellus*, **48A**, 96–121.
- Pu, Z.-X., E. Kalnay, J. Sela, and I. Szunyogh, 1997: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Wea. Rev.*, **125**, 2479–2503.
- , S. Lord, and E. Kalnay, 1998: Forecast sensitivity with dropwindsonde data and targeted observations. *Tellus*, **50A**, 391–410.
- Ralph, F. M., and Coauthors, 1999: The California Land Falling Jets Experiment (CALJET): Objectives and design of a coastal atmosphere–ocean observing system deployed during a strong El Niño. Preprints, *Third Symp. on Integrated Observing Systems*, Dallas, TX, Amer. Meteor. Soc., 78–81.
- Rao, C. R., 1973: *Linear Statistical Inference and Its Applications*. Wiley, 625 pp.
- Szunyogh, I., Z. Toth, K. A. Emanuel, C. H. Bishop, C. Snyder, R. E. Morss, J. Woolen, and T. Marchok, 1999a: Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.*, **125**, 3189–3218.
- , —, S. J. Majumdar, R. E. Morss, C. H. Bishop, and S. Lord, 1999b: Ensemble-based targeted observations during NORPEX. Preprints *Third Symp. on Integrated Observing Systems*, Dallas, TX, Amer. Meteor. Soc., 74–77.
- Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- , I. Szunyogh, S. J. Majumdar, R. E. Morss, B. J. Etherton, C. H. Bishop, and S. Lord, 1999: The 1999 Winter Storm Reconnaissance Program. Preprints, *13th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 27–32.
- , and Coauthors, 2000: Targeted observations at NCEP: Toward an operational implementation. Preprints, *Fourth Symp. on Integrating Observing Systems*, Long Beach, CA, Amer. Meteor. Soc., 186–193.
- Woolen, J. S., 1991: New NMC operational OI quality control. Preprints, *Ninth Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 24–27.
- Zhu, Y., Z. Toth, E. Kalnay, and M. S. Tracton, 1998: Probabilistic quantitative precipitation forecasts based on the NCEP global ensemble. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., J8–J11.