A Regional Model Intercomparison Using a Case of Explosive Oceanic Cyclogenesis

John R. Gyakum,* Marco Carrera,* Da-Lin Zhang,* Steve Miller,[†] James Caveen,[#] Robert Benoit,[#] Thomas Black,[@] Andrea Buzzi,[&] Clément Chouinard,** M. Fantini,[&] C. Folloni,[&] Jack J. Katzfey,[††] Ying-Hwa Kuo,[##] François Lalaurette,[@@] Simon Low-Nam,[##] Jocelyn Mailhot,[#] P. Malguzzi,[&] John L. McGregor,[††] Masaomi Nakamura,[&&] Greg Tripoli,*** and Clive Wilson[†††]

(Manuscript received 9 April 1996, in final form 26 July 1996)

ABSTRACT

The authors evaluate the performance of current regional models in an intercomparison project for a case of explosive secondary marine cyclogenesis occurring during the Canadian Atlantic Storms Project and the Genesis of Atlantic Lows Experiment of 1986. Several systematic errors are found that have been identified in the refereed literature in prior years. There is a high (low) sea level pressure bias and a cold (warm) tropospheric temperature error in the oceanic (continental) regions. Though individual model participants produce central pressures of the secondary cyclone close to the observed during the final stages of its life cycle, systematically weak systems are simulated during the critical early stages of the cyclogenesis. Additionally, the simulations produce an excessively weak (strong) continental anticyclone (cyclone); implications of these errors are discussed in terms of the secondary cyclogenesis. Little relationship between strong performance in predicting the mass field and skill in predicting a measurable amount of precipitation is found. The bias scores in the precipitation study indicate a tendency for all models to overforecast precipitation. Results for the measurable threshold (0.2 mm) indicate the largest gain in precipitation scores results from increasing the horizontal resolution from 100 to 50 km, with a negligible benefit occurring as a consequence of increasing the resolution from 50 to 25 km. The importance of a horizontal resolution increase from 100 to 50 km is also generally shown for the errors in the mass field. However, little improvement in the prediction of the cyclogenesis is found by increasing the horizontal resolution from 50 to 25 km.

1. Introduction

This paper presents a portion of results from an international regional model intercomparison experiment (i.e., Comparison of Mesoscale Prediction and Research Experiments, referred to as COMPARE) using a case of explosive marine cyclogenesis (Chouinard et al. 1994). The purpose of this research is to analyze the results of this first model intercomparison study in an effort to identify important scientific issues relating to the understanding and predictability of mesoscale cyclogenesis. Since such an intercomparison of explosive secondary cyclogenesis is unprecedented, our study will define the state-of-the-art performance of regional models in the simulation of such an event. The focus of our analysis is on the composite forecast results, since one of the objectives of our study is to identify systematic errors in regional modeling of secondary cyclogenesis. This study will evaluate the horizontal and vertical resolutions necessary to make a credible simulation. We will also address the issues of model physics and initial conditions.

A perspective of COMPARE is found in its long-term objectives: 1) to propose and perform model and data assimilation intercomparison experiments in a collaborative and scientifically controlled manner to further understanding and predictive capability at the mesoscale; 2) to identify important issues of mesoscale research and prediction that may be addressed by numerical experimentation; and 3) to establish over a period of years a test bed of a broad range of mesoscale

* Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada.
[†] Maritimes Weather Centre, Atmospheric Environment Service (AES), Bedford, Nova Scotia, Canada.
[#] Recherche en Prévision Numérique, AES, Dorval, Quebec, Canada.
[@] National Centers for Environmental Prediction, Camp Springs, Maryland.
[&] National Research Council of Italy, FISBAT Institute, Bologna, Italy.
** Data Assimilation and Satellite Meteorology Division, AES, Dorval, Quebec, Canada.
[††] Commonwealth Scientific and Industrial Research Organisation, Victoria, Australia.
[##] National Center for Atmospheric Research, Boulder, Colorado.
[@@] Météo-France, Toulouse, France.
[&&] Japan Meteorological Agency, Tokyo, Japan.
*** Department of Atmospheric and Oceanic Sciences, University of Wisconsin—Madison, Madison, Wisconsin.
[†††] U. K. Meteorological Office, Bracknell, United Kingdom.

Corresponding author address: Dr. John R. Gyakum, Department of Atmospheric and Oceanic Sciences, McGill University, 805 Sherbrook Street West, Montreal PQ H3A 2K6, Canada.
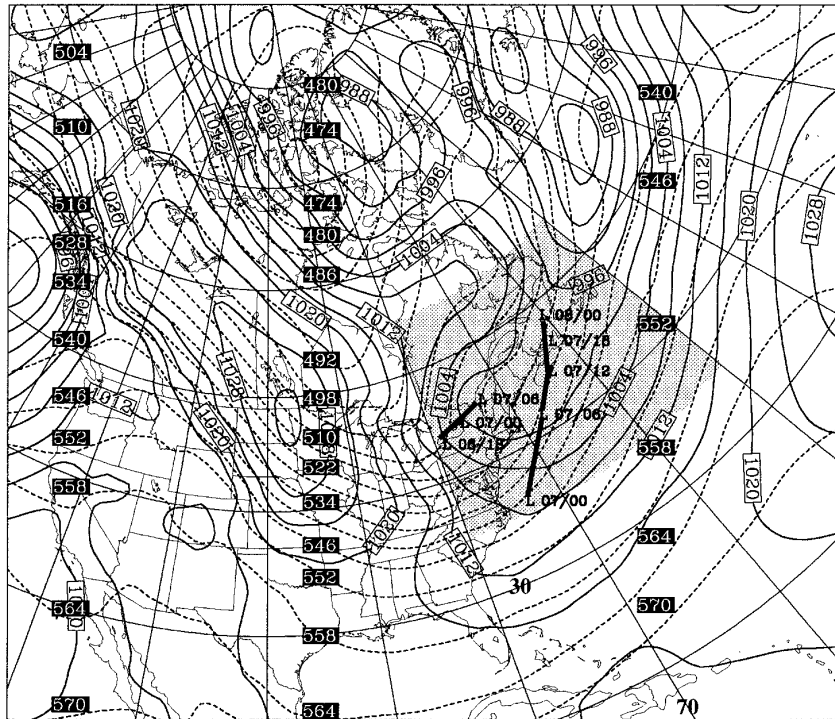E-mail: gyakum@zephyr.meteo.mcgill.ca

FIG. 1. Time-mean SLP (solid, at intervals of 4 hPa) and 1000–500-hPa thickness (dashed, at intervals of 6 dam) for the period 1200 UTC 6 March–0000 UTC 8 March 1986. The verification domain is shown by the shaded region. Analyzed primary and secondary cyclone tracks are shown with date/time. Latitude–longitude lines are shown each 10°, as is the case for subsequent maps.

cases using high quality raw datasets, assimilation systems, and analyses selected primarily from intensive observation periods (IOPs) of well-instrumented observational campaigns.

The cyclogenesis event, being used for COMPARE, occurred during the concurrent Canadian Atlantic Storms Program (CASP; Stewart et al. 1987) and the Genesis of Atlantic Lows Experiment (GALE; Dirks et al. 1988) during March 1986. This event has been studied observationally by Yau and Jean (1989) and Stewart and Donaldson (1989), and numerically by Mailhot and Chouinard (1989). Since this case began in the preexisting cyclonic circulation of a large-scale surface cyclone, we will refer to this cyclogenesis as a secondary development. This particular type of cyclogenesis, even when occurring over the data-rich continent, represents an especially difficult scientific forecast challenge (Kuo et al. 1995). This challenge is being met with the efforts of the COMPARE participants, representing 10 different modeling groups in seven countries.

The outline of the paper includes the COMPARE experimental design in section 2, the evaluation experimental design in section 3, a synoptic intercomparison of surface systems in section 4, an evaluation of rms errors and bias in section 5, a discussion of S1 scores

and cyclone structure in section 6, a discussion of potential vorticity in section 7, precipitation verification in section 8, and the conclusions in section 9.

2. Experimental design for COMPARE

The oceanic cyclogenesis case chosen for COMPARE is one occurring during the 14th intensive observing period (CASP IOP 14) from 1200 UTC 6 March 1986 through 0000 UTC 8 March 1986. Consistent with COMPARE objective 3, this first case is documented with additional buoy, ship, radiosonde, and aircraft dropsonde data to supplement the conventional surface and radiosonde reports. Additionally, coastal radiosonde stations in the United States and Canada took 6-hourly observations.

A large-scale perspective is provided by the 36-h time-averaged sea level pressure (SLP) and 1000–500-hPa thickness (Fig. 1), based upon the analyses of Chouinard et al. 1994, with the smaller verification domain (shaded) for this simulation. The case is characterized by cyclonic flow at lower levels with a strong southwesterly thermal wind to the east of a trough in the Great Lakes region. A quasi-stationary surface anticyclone gradually weakens through the period in the cold air in the northern region of the verification do-

main, while two surface cyclones are active to the south (see tracks in Fig. 1). The southernmost system that travels from the U.S. east coast into the Canadian province of Nova Scotia is the focus of our evaluation. However, we will also examine the simulations of the larger-scale anticyclone and primary cyclone in the context of some of the scientific objectives discussed by Chouinard et al. (1994). These objectives include upper-level dynamical support, coastal front enhancement, the roles of sensible and latent heat fluxes and the low-level jet, and latent heat release in frontal clouds.

The analysis procedure involves the use of the Canadian Regional Data Assimilation System (Chouinard et al. 1994) to generate 6-hourly fields at a horizontal resolution of 50 km and a vertical resolution of 25 hPa. Participants using polar stereographic projection domains were provided with these analyses for the full region of Fig. 1. Integrations were initialized with the analysis at 1200 UTC 6 March 1986. The lateral boundary conditions are derived from the 6-h analyses. As pointed out by Chouinard et al. (1994), the larger grid of Fig. 1 is sufficiently large (compared with the shaded region) so that the information at the boundaries does not have time to reach the shaded verification region and artificially improve the forecast quality (e.g., Fig. 17 of Chouinard et al. 1994). For participants using models with domains other than polar stereographic, a choice of slightly larger latitude–longitude domains was provided (see Fig. 10 of Chouinard et al. 1994).

Both the analyses and radiosonde observations, within the shaded region of Fig. 1, are used to verify the six experiments (Table 1). The first two involve a test of varying the vertical resolution (18 and 35 levels), while maintaining a 100-km horizontal resolution. Experiments 3 and 4, with 50-km resolution, include the same vertical resolution test. Experiment 5 is the high-resolution experiment in which 52 levels are used on a horizontal grid of 25-km resolution. The final experiment 6 consists of only a 24-h simulation that is initialized 12 h later than the other experiments. It is designed to test the spinup of numerical models being used. All participants were supplied with the same initial conditions, either on a polar stereographic domain (true at 60°N) at 50-km resolution, or a latitude–longitude domain at 0.5° resolution, depending on their model grid structure. Only suggestions for the thickness of each vertical layer and the horizontal mesh configuration were given to the participants. As they were not mandatory, some groups made choices that were significantly different. The initial time is 1200 UTC 6 March 1986 in the first five experiments. All datasets generated by these groups are interpolated back to our verification grid mesh for evaluation.

The participating institutions and models are listed in Table 2. They include the CSIRO (Commonwealth Scientific and Industrial Research Organisation) Limited Area Model (LAM; McGregor 1993), AES (Atmospheric Environment Service) Recherche en Prévision Numérique (RPN) Regional Finite Element Model (RFE; Tanguay et al. 1989; Benoit et al. 1989; Mailhot et al. 1995), AES nonhydrostatic Mesoscale Compressible Community Model (MC2; Tanguay et al. 1990; Benoit et al. 1996), Météo-France PERIDOT model (Prévisions à Echéance Rapprochée Intégrant des Données Observées et Télédétectées; Imbard et al. 1987), the Limited Area Model (BOLAM) of Italy's National Research Council Institute of Physics and Chemistry of the Low and High Atmosphere (FISBAT Institute; Buzzi et al. 1994), the Japanese Meteorological Agency (JMA) Japan Limited Area Spectral Model (JLASM; Segami et al. 1989), United Kingdom Meteorological Office (UKMO) model (Cullen and Davies 1991), the Pennsylvania State University–Na-

TABLE 1. List of experiments with numbers of participants (letter identifier is shown).

Experiment		1	2	3	4	5	6	
Horizontal resolution (km)		100	100	50	50	25	100	
Levels		18	35	18	35	52	18	
Institution model	Letter identifier							Total
CSIRO LAM	C	1	1	1	1	1	1	6
AES RFE	R	1	1	1	1	1	1	6
AES MC2	M	1	1	1	1	1	0	5
Météo-France, Toulouse	T	1	1	1	1	0	1	5
FISBAT, Italy	B	1	1	1	1	0	1	5
JMA JLASM	J	1	1	1	0	0	0	3
UKMO	E	1	1	1	1	0	1	5
	K	1	1	1	1	1	1	6
	A	1	1	1	1	1	1	6
PSU–NCAR MM4	P	1	1	1	1	0	0	4
NCEP eta	N	1	1	1	1	0	0	4
UW—NMS	W	1	0	1	0	0	0	2
Total number		12	11	12	10	5	7	57

TABLE 2. Summary of model characteristics.

Institution model	PBL	Deep convection
CSIRO LAM	K theory function of Richardson number (Ri), referred to as K (Ri)	Arakawa (1972)
AES RFE	Turbulent kinetic energy (TKE), Benoit et al. (1989)	Kuo (1974)
AES MC2	TKE, Benoit et al. (1989)	Kuo (1974)
Météo-France PERIDOT	TKE, Bougeault and Lacarrère (1989)	Mass flux
FISBAT, Italy	K (Ri)	Emanuel (1991)
JMA JLASM	Mellor and Yamada (1974)	Convective adjustment (Gadd and Kears 1970)
UKMO	K (Ri) for all	Gregory and Rowntree (1990) for all
1		
2		
3		
PSU–NCAR MM4	Blackadar scheme, Zhang and Anthes (1982)	Grell (1993)
NCEP eta	Mellor and Yamada (1974)	Betts and Miller (1986)
UW–NMS	K theory (horizontal), TKE (vertical)	Kuo (1974)

tional Center for Atmospheric Research (PSU–NCAR) Mesoscale Model Version 4 (MM4; Anthes et al. 1987), the National Centers for Environmental Prediction's (NCEP, formerly the National Meteorological Center) step-mountain eta model (Mesinger et al. 1988), and the University of Wisconsin—Nonhydrostatic Modeling System (UW—NMS; Tripoli 1992). These models are integrated with their state-of-the-art physical representations for air–sea interaction, boundary layer physics, and convective diabatic heating. All models, except for the AES MC2 and the UW—NMS, are hydrostatic.

The extent of participation varied substantially among the experiments and ranged from 5 for the highest-resolution experiment 5 to 14 for the low-resolution experiment 1. Substantial flexibility was offered to the participants; each modeling group was free to provide as few or as many simulations as desired. For example, the UKMO produced three runs for five of the experiments to test the sensitivity to different surface roughness representations. The first (E in Table 1) uses roughness lengths that are vegetation dependent and typically small (<0.1 m), the second (K) includes an additional component associated with orography, and the third (A) includes a new surface parameterization that takes into account the form drag due to orography, taking into account the ideas of Wood and Mason (1993). This latter scheme leads to effective roughness lengths for momentum up to 50 m in mountainous terrain. A total of 57 simulations is evaluated for this study.

3. Evaluation methodology

Our evaluation begins with a presentation of central pressure ensembles, or groupings, of individual model outputs for each circulation system. Domain-averaged scoring consists of the rms error and bias, respectively; the average model forecast error; and the difference between the forecast and analyzed fields. Additionally, we use the S1 score (Teweles and Wobus 1954), which measures the errors in horizontal gradients.

The parameters are defined as

$$\text{rms error} = \left[\frac{1}{N} \sum (F_i - A_i)^2 \right]^{0.5} \quad (1)$$

and

$$\text{bias} = \frac{1}{N} \sum (F_i - A_i), \quad (2)$$

where N is the total number of grid points, summed (Σ) from the first to the final ith point, and F and A are the forecasted and analyzed variables, respectively. The S1 score is

$$\text{S1} = 100 \frac{\Sigma |e_G|}{\Sigma |G_L|}, \quad (3)$$

where e_G is the error of the forecasted pressure difference and G_L is the maximum of either the forecasted or analyzed pressure difference between two grid points.

Systematic evaluation of surface cyclone tracks and composite (or the mean of the simulations) errors in sea level pressure, layer thickness, and potential vorticity will also be discussed. The continuity of individual cyclone centers is documented by tracking closed geostrophic circulation centers. When a closed circulation does not exist, as is the case at formation time of the secondary low, the geostrophic vorticity maximum is used as its position. The role of resolution is examined as follows: Experiment 3 (E3) is used as a control simulation, from which we may decrease the horizontal resolution to 100 km (E1), or increase the vertical resolution to 35 levels (E4).

4. Surface feature intercomparison and associated systematic errors

The ensembles of central pressure for the secondary cyclogenesis are shown in Fig. 2. Experiment E1 yields

substantial variability, especially at hours 30 and 36 (Fig. 2a). The RPN-analyzed intensification rate, at 1.3 Bergerons[1], qualifies the system as explosively intensifying. It is encouraging that several of the models with such a resolution (100 km/18 levels) captured this intensification, which illustrates the dramatic progress in simulating this phenomenon since Sanders and Gyakum (1980) found its systematic underprediction. Results from E3 (Fig. 2b), with 50-km resolution and the same vertical resolution, show similar results, though with more participants exceeding the analyzed intensity during the final 6 h. Since the participation varied substantially among the experiments, we assess the sensitivity of the results with composites for each experiment (Fig. 2c), derived from the four common models (C, R, K, and A). *Clearly, the most substantial increase (among the sensitivity experiments) in cyclogenesis occurs with the enhancement of horizontal resolution from 100 to 50 km.* There is only a slight change as a result of enhancing the vertical resolution at either 50- or 100-km resolution, and there is little improvement as a result of enhancing the horizontal resolution to 25 from 50 km. The later initialization experiment E6 shows a systematic overprediction during the 30–36-h period.

The secondary cyclone tracks of the E3 cyclone centers (Fig. 3) show considerable variability. However, *there is a systematic tendency for most of the models to predict an excessively westward position for this cyclone.* This result holds for all experiments (not shown) and was also found by Oravec and Grumm (1993, their Fig. 5) to be the case in the operational NCEP Nested Grid Model (NGM) during the winter of 1991. This error is associated with a slow bias in the speed of the secondary system, in which the participants' mean error is 150 km to the west and south of the observed system by 0000 UTC 8 March. The precise reasons for this systematic error are beyond the scope of this work. However, we suggest later an association with excessive lower-tropospheric stabilization over the track of the cyclone. Such a systematic error would produce slower movement of the low (e.g., Bluestein 1993, 46–47). The primary cyclone tracks (Fig. 3), originating near Lake Ontario, show a systematic northeastward curvature, which, as we will see later, is associated with the excessive simulated intensity.

At 0600 UTC 7 March, the RPN-prepared COMPARE analysis shows three separate surface cyclones (Fig. 4), with 988-hPa central pressure in the two maritime systems. The most northwest system is the primary low (989 hPa) and loses its identity thereafter. Figure 5, showing the central pressure traces of the analysis and the participating models in E3, reveals the

---

[1] Bergeron is a geostrophically equivalent rate to a central pressure fall of 1 mb h^{-1} for 24 h at 60°N, see Sanders and Gyakum 1980)
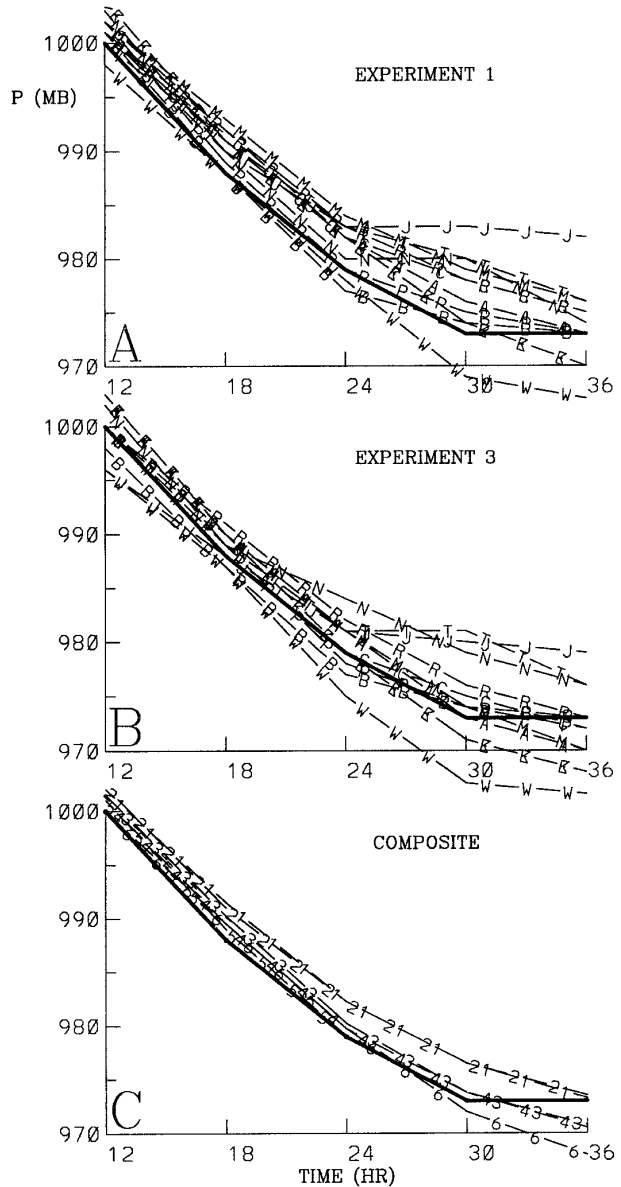

FIG. 2. Ensembles of secondary cyclone central pressure (hPa) as a function of time (h), where 12 h corresponds to 0000 UTC 7 March 1986 for (a) experiment 1 (100 km/18 levels) and (b) experiment 3 (50 km/18 levels). Composite central pressures for each experiment (indicated as numbers from 1 to 6 corresponding to the listing in Table 1), derived from the four models participating in all, are shown in panel (c). For E6, initialized 12 h later than the other simulations, 12 h corresponds to the initial time at 0000 UTC 7 March. The RPN-prepared COMPARE (solid) analyses is shown in all panels. Letters used in the pressure traces identify each model (see Table 2). Such a convention is used for the remaining figures.

simulations to overdeepen the primary system. By 0600 UTC 7 March (at 18 h), all except one of the models show excessively low pressure at its center. Thereafter, 10 of the 12 participants continue to deepen the system during a time when the real system loses its closed
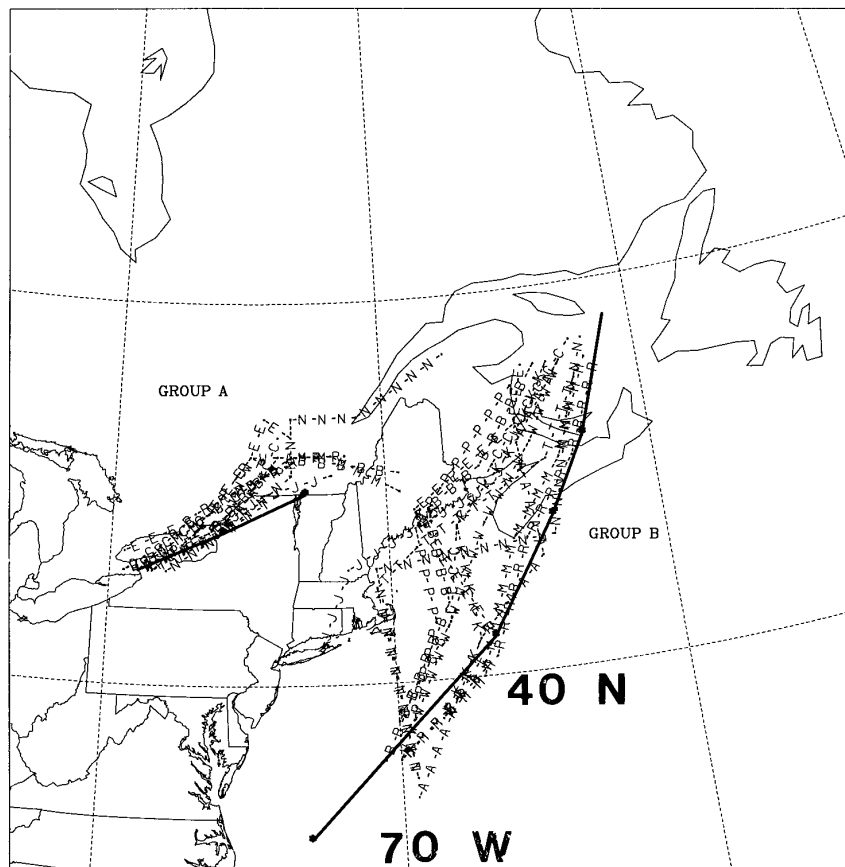
FIG. 3. Tracks of the primary (group A) and secondary (group B) cyclone centers for each of the 12 E3 participants (based upon the 6-h positions) from 0000 UTC 7 March through 0000 UTC 8 March 1986 and (b) the primary cyclone center beginning at 1800 UTC 6 March 1986. The verifying RPN-prepared COMPARE analyses are shown by the solid lines.

circulation. Leary (1971) and Silberberg and Bosart (1982) identified more systematically that operational models overdeepened continental cyclones while producing weaker-than-observed oceanic cyclones. More recently, Junker et al. (1989) have shown the NCEP NGM to produce systematically weak (strong) cyclones over the ocean (land) regions. As pointed out earlier in reference to Fig. 2, the models performed more credibly in predicting the secondary cyclogenesis, although there is a slight systematic tendency to underpredict the system early in its life cycle.

To reinforce our previous conclusions in reference to both cyclones' central pressures, we present Fig. 6, which shows the difference between the composite E3 SLP and the corresponding analysis at 0600 UTC 7 March 1986. Clearly, *there is a systematic SLP excess over the maritime regions and a deficit over the land area.* The maximum error in the vicinity of the secondary low (Fig. 4) exceeds 4 hPa, whereas near the primary low it is as low as −5 hPa. The offshore pressure excess is sufficiently strong to dominate the domain-averaged statistics, to be discussed later. The off-

shore bias increases to a maximum of 6.5 hPa when the horizontal resolution decreases to 100 km at 18 levels (i.e., from E3 to E1), and the inshore bias is minimized to −1.6 hPa when the models were initialized at a later time in E6 (not shown).

Associated with SLP errors are tropospheric temperature errors (Fig. 7): We display the E3 1000–500-hPa thickness composite minus the corresponding analysis at 0600 UTC 7 March 1986 (Fig. 7). There is a predominant maritime/coastal (continental) tropospheric cold (warm) bias, qualitatively similar to that found by Junker et al. (1989) for the NCEP models. The cold bias is most prominent near the secondary low, and the warm bias is maximized near the primary low (Fig. 4). These results are qualitatively the same as those of Leary (1971), in which the oceanic cyclones are too weak and cold and the continental systems are too strong and warm. The offshore cold bias is largest in the lower resolution runs at 100 km and 18 levels. However, there is some improvement in the land warm bias from initializing 12 h later in E6 (not shown).
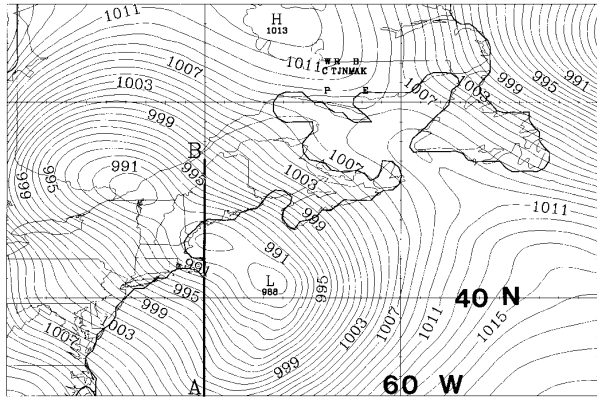
FIG. 4. Sea level pressure field (at intervals of 1 hPa) analysis at 18 h (0600 UTC 7 March 1986). Cross section used in Fig. 9 is shown in thick solid lines (A–B). Anticyclone positions from each of the E3 participants (letter coded as in other figures) are also shown. The land–ocean mask used for Fig. 16 is also shown. Latitude–longitude lines are shown each 10°.
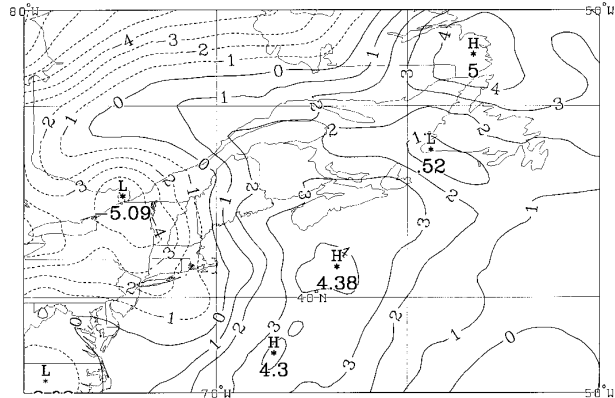

FIG. 6. Difference of E3 (50 km/18 levels) SLP composite and the RPN-prepared COMPARE analysis (interval of 1 hPa, with dashed showing negative) at 0600 UTC 7 March 1986.

Figure 4 also shows the predicted positions of a surface anticyclone by the E3 participants in relation to the observed 1013-hPa system in the extreme north-central part of the domain. The systematic tendency of the models to locate the system too far to the southeast (see the clustering of the lettered simulations in Fig. 4) is evident. Additionally, the central pressures of the system are too low among many of the participants in E3 after 12 h (Fig. 8). This weaker-than-observed system is also too warm (Fig. 7). Grumm and Gyakum (1986) and Junker et al. (1989) have found similar

systematic errors in their study of operational NCEP models for anticyclones in this region.

The north–south cross section along 35°–47°N (A–B shown in Fig. 4) of the analyzed potential temperature through the U.S. coastal region (Fig. 9a) illustrates two baroclinic zones: One along the secondary cyclone track between 38° and 41°N (Fig. 3), and the other in the geostrophic southeasterlies between the surface anticyclone and the primary low (Fig. 4). As would be expected, the stratification is relatively strong in the vicinity of the continental anticyclone and much weaker along the oceanic secondary storm track. The E3 composite errors in potential temperature (Fig. 9b) reveal that i) the marine environment south of 40°N is too cold in the troposphere and is excessively stable statically in the lower troposphere, ii) the lower troposphere north of 44°N in the vicinity of the anticyclone (Fig. 4) is too warm and unstable, and iii) these errors are associated with excessively weak lower-tropospheric baroclinity.

FIG. 5. Ensembles of the primary cyclone's central pressure (hPa) as a function of time (h), where 6 h corresponds to 1800 UTC 6 March 1986, for experiment 3. The RPN-prepared COMPARE (solid) analysis is shown.
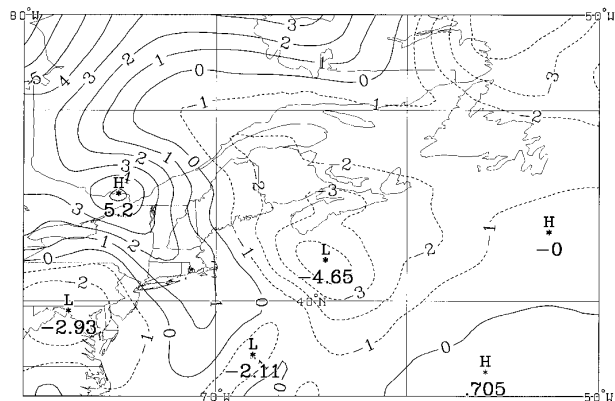

FIG. 7. Difference of E3 1000–500-hPa thickness composite and the RPN-prepared COMPARE analysis (interval of 1 dam, with dashed showing negative) at 0600 UTC 7 March 1986.
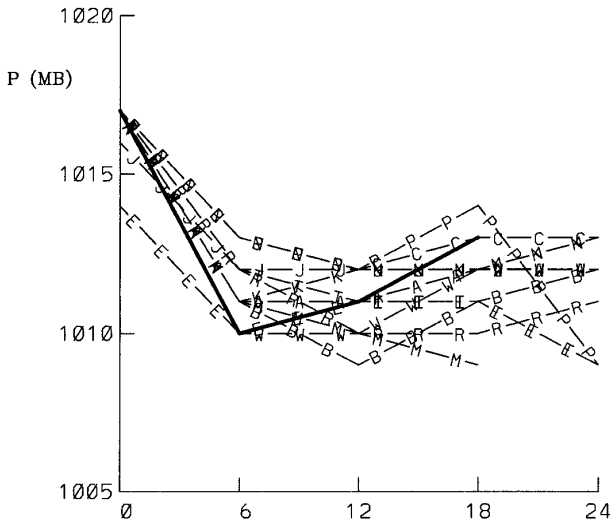
FIG. 8. Central pressures of surface anticyclone discussed in the text, as a function of hour, for the participants in E3, with the RPN-prepared COMPARE (solid) verifying analysis.

5. Rms errors and bias

Since E3 corresponds closely to the analyzed secondary cyclone's central pressure (Fig. 2c) with a large number (12) of participants, we focus on traditional scores of rms error and bias for this run at 24 h (1200 UTC 7 March), when the cyclogenesis is well under way (Fig. 2). Figure 10 shows the vertical structure of wind speed, temperature, and height rms error. The speed rms errors peak near 800 and 300 hPa, where the respective lower- and upper-level jets are located. A qualitatively similar structure is seen in the temperature errors, but the peaks occur at more elevated levels between 200 and 250 hPa and in the lowest kilometer of the planetary boundary layer. The composite height rms error of 25–30 m varies weakly with height. A direct quantitative comparison with earlier studies on such errors is not possible, since we are examining one case on a particular grid and domain that may be very different from those evaluated in past studies. Nevertheless, the rms height errors found here appear considerably improved over the 30–45-m values published by Anthes (1983) for 24-h forecasts. The temperature rms error is improved by about 1°C from this 1980–82 period studied by Anthes. Biases of the same fields (Fig. 11) show systematic underestimates of wind speeds, with the composite approximately -2 m s^{-1} at the levels of the lower and upper troposphere. Substantial scatter exists in the temperature bias distribution (Fig. 11b) with some participants showing cold biases and others warm. Most participants have a positive height bias throughout the troposphere (Fig. 11c).

The time series of 300-hPa rms wind speed errors shows an especially rapid growth during the first 12 h of integration, with evidence for a decline during the

final 6 h (Fig. 12a). The large error growth occurs during the period leading up to, and during, the most rapid secondary cyclogenesis (Fig. 2). The 850-hPa temperature rms error (Fig. 12b), in contrast, also grows with time more rapidly, but later in the forecast period, showing large variability among the participants. This large variance may be due in part to the substantial difference in secondary cyclone track (to be discussed in the next section). Additionally, errors appear to peak at the hours of 6, 18, and 30—all times when there is less sounding coverage and presumably less reliable analyses. The secondary cyclone's intensity reaches its peak at 30 h (Fig. 2c), and this forecast challenge may be contributing a maximum in error at this time. Like
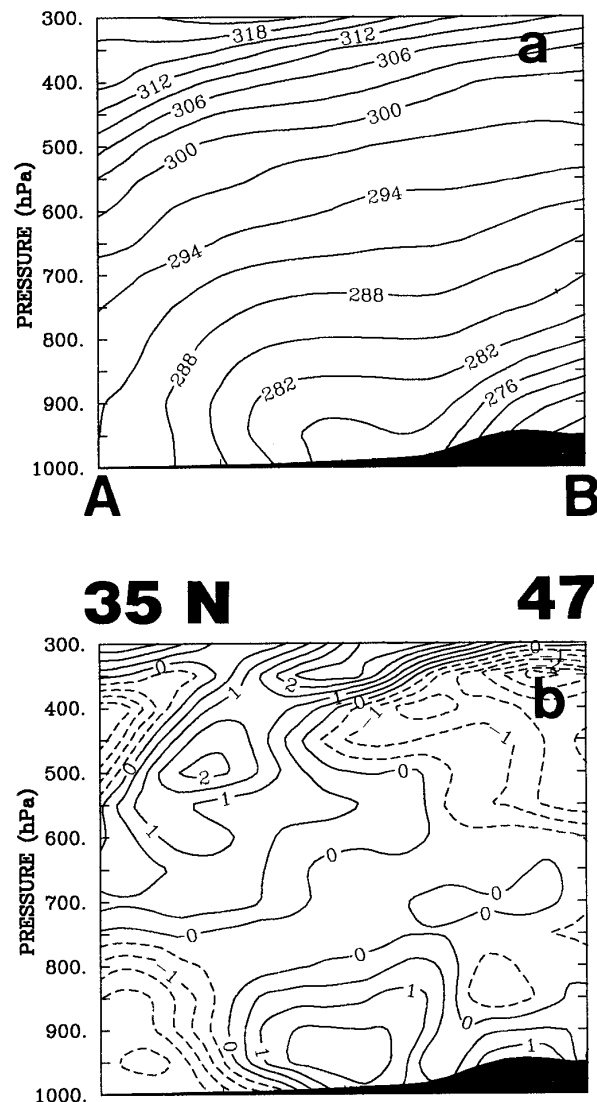

FIG. 9. Cross section (shown in Fig. 4) of (a) analyzed potential temperature (interval of 3 K) and (b) E3 composite potential temperature minus analyzed (interval of 0.5 K, with dashed corresponding to negative).
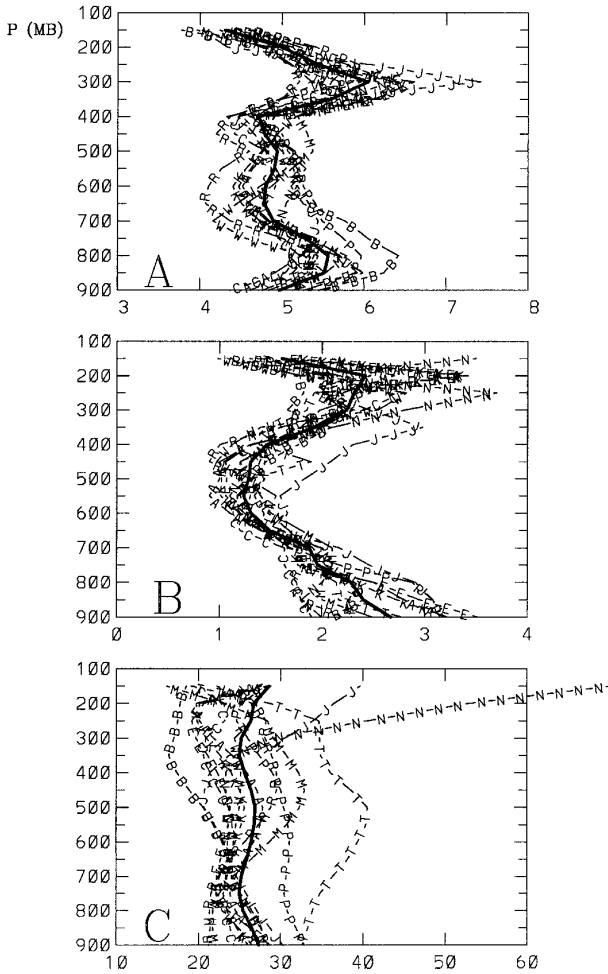
FIG. 10. Vertical distribution of rms errors, with respect to the RPN-prepared COMPARE analysis, in (a) wind speed (m s^{-1}), (b) temperature (°C), and (c) height (m) for E3 (50 km/18 levels) at 24 h with the 12 participants (lettered) and composites (solid) shown.

the temperatures, the height at 300 hPa (Fig. 12c) shows a similar structure in error growth. The errors at the initial time are all nonzero. This is due primarily to interpolation errors of the model fields onto the verifying grid.

The time series in 300-hPa wind speed bias shows an evolution to a slow bias by 12 h with the composite showing approximately a 1 m s^{-1} deficit (Fig. 13). Though there is substantial rms error by 6 h (Fig. 12a), there is little bias. However, as has been pointed out earlier, the reliability of the scoring at 6, 12, and 18 h is questionable. A systematic cold bias (Fig. 13b) of about 1°C with large variance later in the forecast period is evident. This bias is primarily occurring over the marine area of the verification domain (shaded region of Fig. 1). Finally, though large rms errors are present in the 300-hPa geopotential (Fig. 12c), there

is little composite bias (Fig. 13c) in the field, even though most participants show a positive bias.

The composite rms and bias errors at 1200 UTC 7 March, respectively shown in Figs. 14 and 15, reinforce our earlier conclusion, based upon central pressure, that *the most substantive benefit is derived from increasing the horizontal resolution from 100 to 50 km with 18 levels* (i.e., from E1 to E3). Since the physical parameterizations behave very differently at these different resolutions, this reduction in error from the resolution increase is likely to be a consequence of several factors. The only exception to the general improvements occurring primarily from the horizontal resolution increase is in the geopotential heights (Figs. 14c and 15c) in which there is some systematic improvement in increasing the number of vertical levels, while maintaining a 50-km horizontal resolution (i.e., from E3 to E4). *A particularly intriguing result is that negligible improvement occurs as a consequence of increasing both the horizontal and vertical resolutions to 25 km and to 52 levels (from E4 to E5).* This is further evidence that increasing the model resolution does not necessarily increase skill.
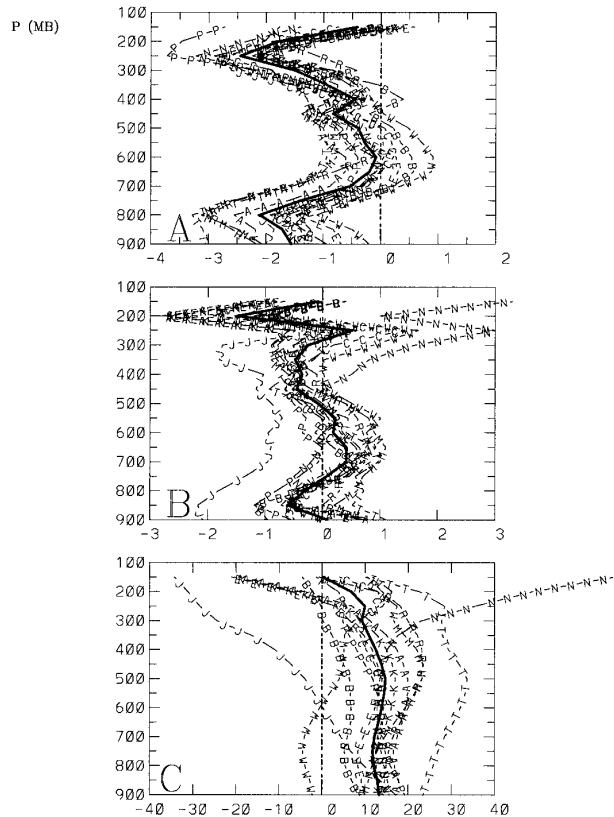

FIG. 11. Vertical structures of bias, with respect to the RPN-prepared COMPARE analysis, for (a) wind speed (m s^{-1}), (b) temperature (°C), and (c) height (m) for E3 (50 km/18 levels) at 24 h with the 12 participants (lettered) and composites (solid) shown.
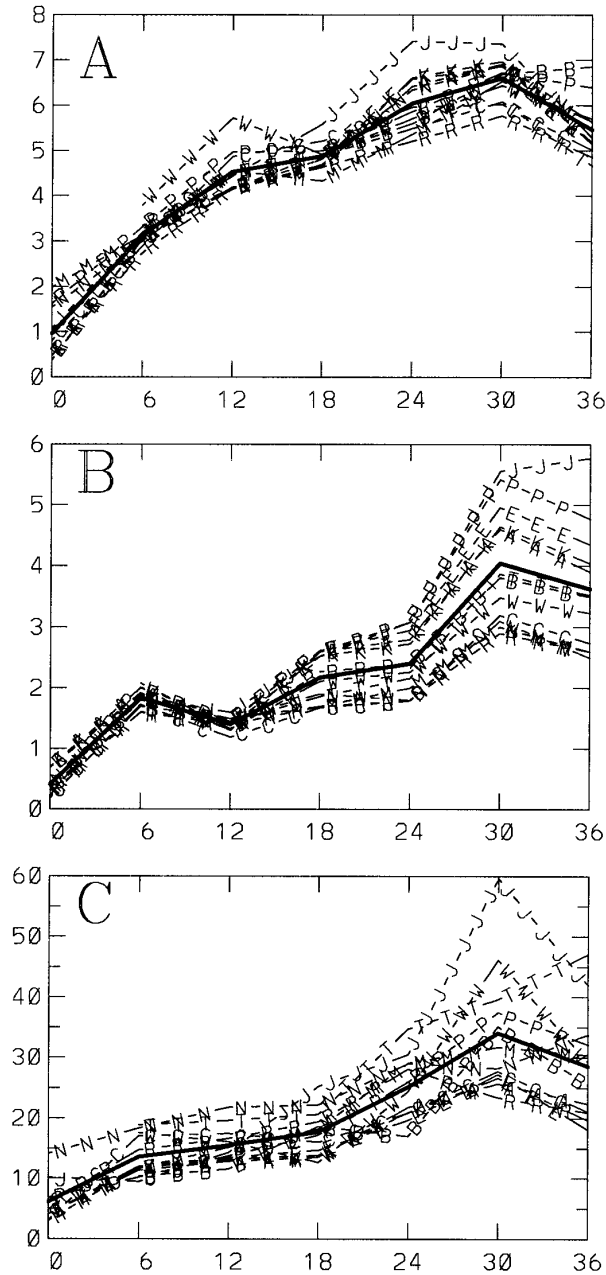
FIG. 12. Time series of rms errors, with respect to the RPN-prepared COMPARE analysis, for (a) 300-hPa wind speed (m s^{-1}), (b) 850-hPa temperature (°C), and (c) 300-hPa height (m) for E3 (50 km/18 levels) with the 12 participants (lettered) and composites (solid) shown. Data for the participants W, B, and T at the initial time are not available.

ing verification. Table 4 shows a compilation of maximum ("worst") and minimum ("best") scores for each participant, as a function of range and experiment. Each participant has at least three best scores (defined as a minimum in either the rms error or in the magnitude of the bias). The fact that each participant excels in simulating a specific field, such as height, wind, or
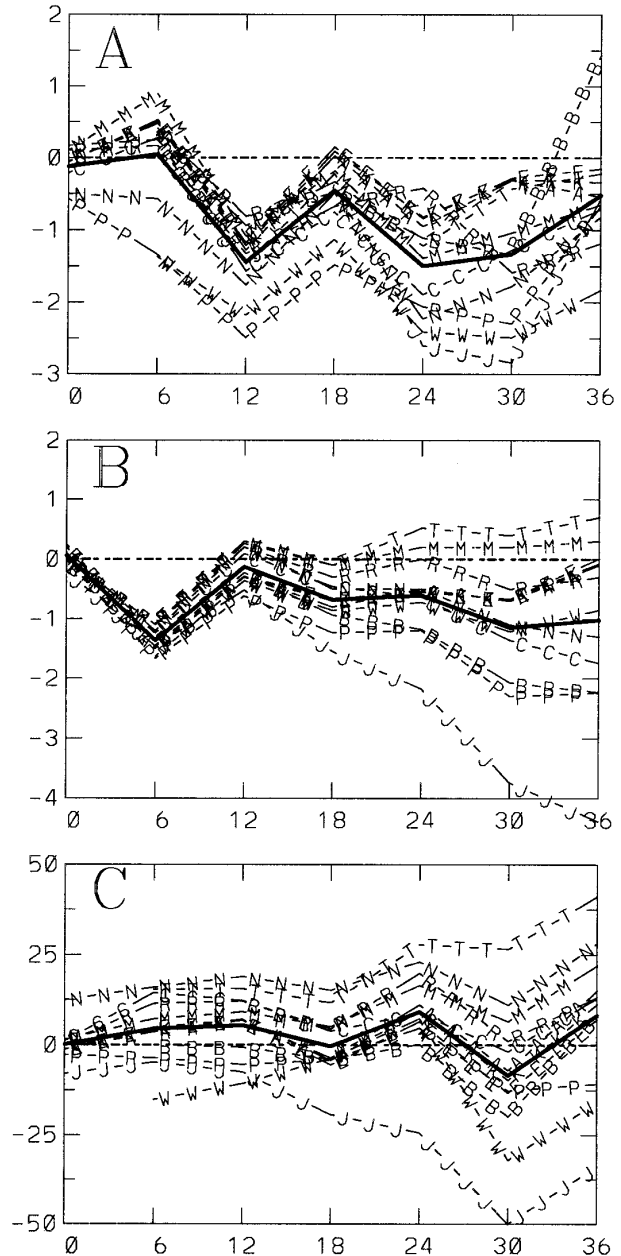

FIG. 13. Time series of bias, with respect to the RPN-prepared COMPARE analysis, for (a) 300-hPa wind speed (m s^{-1}), (b) 850-hPa temperature (°C), and (c) 300-hPa height (m) for E3 (50 km/18 levels) with the 12 participants (lettered) and composites (solid) shown. Data for the participants W, B, and T at the initial time are not available.

To summarize the performance among the participants, we tabulate basic scoring parameters for E1 and E3, as listed in Table 3. These two experiments were chosen since 12 common participants could be identified for each. These standard parameters of rms error and bias are used for both the analysis grid and sounding
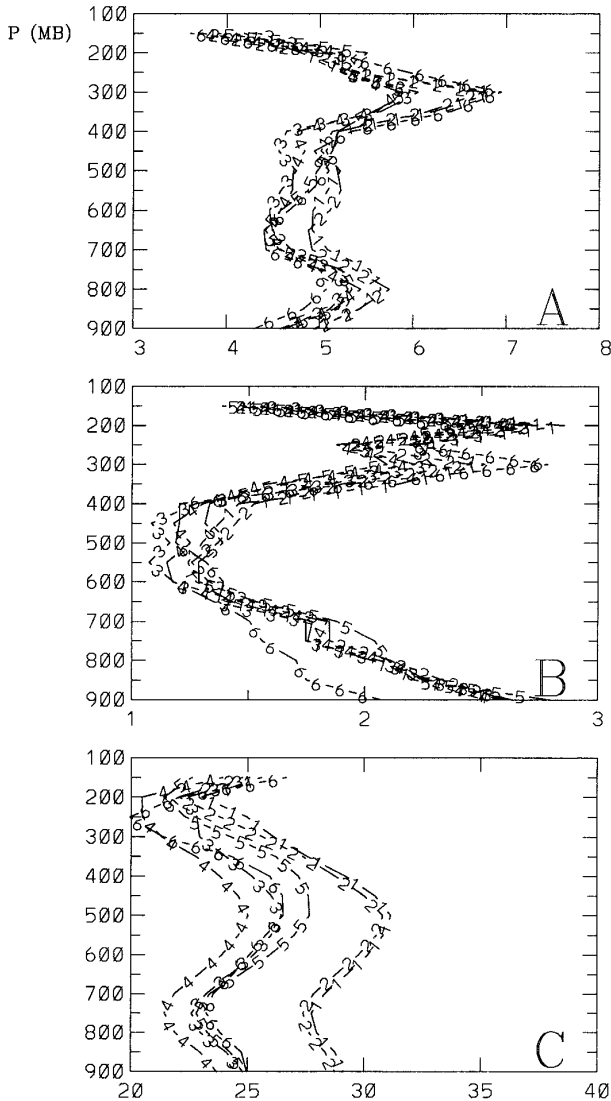
FIG. 14. Vertical profiles of rms errors, with respect to the RPN-prepared COMPARE analysis, composited for each experiment, and derived from the four models participating in all, for (a) wind speed (m s^{-1}), (b) temperature (°C), and (c) geopotential height (m).
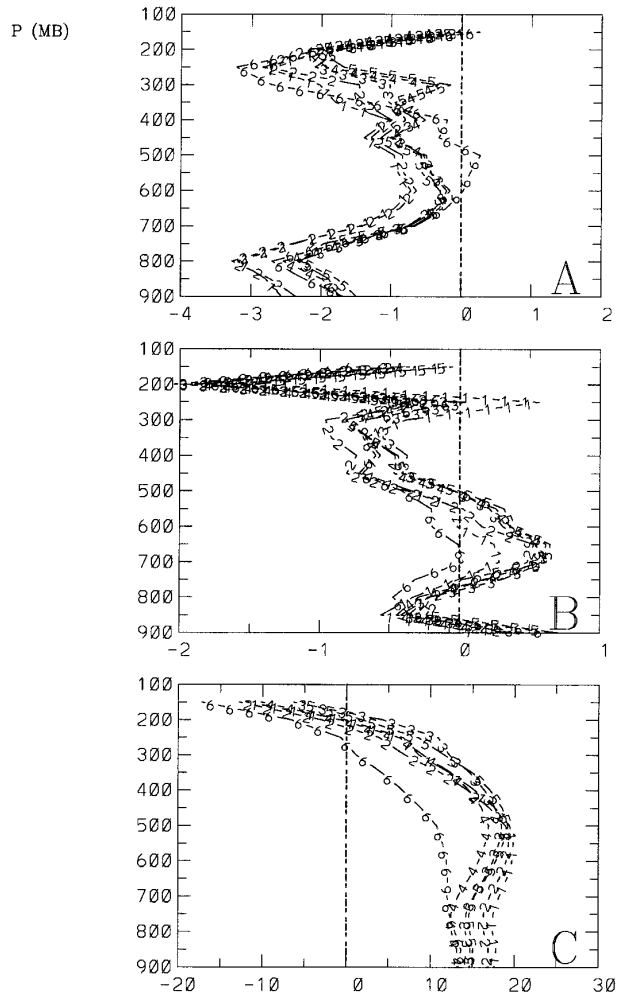

FIG. 15. Vertical profiles of bias, with respect to the RPN-prepared COMPARE analysis, composited for each experiment and derived from the four models participating in all, for (a) wind speed (m s^{-1}), (b) temperature (°C), and (c) geopotential height (m).

temperature, makes impossible the task of finding a specific model and physics package that is ideal for simulating this case of secondary cyclogenesis. A closer examination of Table 4 might tempt the reader to conclude that the RFE is the best performer for this case with 19 leading scores. However, this model's analysis is used as the verifying grid for the participants. Since the verifying grid (Fig. 4) includes a large region of the ocean that lacks the dense upper-air network of the continent, we perform the same exercise of scoring the participants directly against the sounding data (Table 5). By using this sounding set as the "ground truth," the RFE's leading score number drops to 9. Though five other participants appear to benefit

from scoring directly against the soundings (e.g., the UKMO's K simulations increasing from 4 to 13 leading scores), the others fared slightly worse than, or the same as, in the full-gridded scoring scheme. When we

TABLE 3. Parameters used in scoring extremes for Tables 4 and 5.

Variable	Scoring parameter	
	Rms error	Bias
300-hPa temperature	X	X
300-hPa height	X	X
850-hPa temperature	X	X
850-hPa height	X	X
300-hPa wind speed	X	X
850-hPa wind speed	X	X

WEATHER AND FORECASTING

TABLE 4. Frequency of maximum/minimum scores from the analysis grid (parameters listed in Table 3).

Forecast range (h)	12	12	24	24	36	36	
Experiment	1	3	1	3	1	3	
Horizontal resolution (km)	100	50	100	50	100	50	
Institution model (identifier)							Total
CSIRO LAM (C)	0/2	0/2	1/1	0/1	0/0	0/0	1/6
AES RFE (R)	0/3	0/1	0/5	0/4	0/4	0/2	0/19
AES MC2 (M)	1/0	1/0	0/1	0/0	0/2	0/1	2/4
Météo-France, Toulouse (T)	1/0	0/0	5/0	5/1	2/1	2/1	15/3
FISBAT, Italy (B)	0/2	0/4	1/2	1/3	1/0	1/0	4/11
JMA JLASM (J)	1/2	1/1	2/0	5/0	2/0	2/1	13/4
UKMO							
(E)	0/0	0/0	0/0	0/0	0/0	0/4	0/4
(K)	0/0	0/2	0/0	0/0	0/2	0/0	0/4
(A)	0/2	0/1	1/0	0/1	0/3	0/1	1/8
PSU–NCAR							
MM4 (P)	2/0	2/0	1/1	1/0	1/1	1/1	8/3
NCEP eta (N)	4/1	4/2	1/0	1/0	2/2	2/1	14/6
UW—NMS (W)	4/0	4/0	0/3	0/2	4/0	5/0	17/5

consider the verification against the soundings, each participant had at least some (two) ''winning'' scores.

To address more quantitatively the issue of whether scoring the participants against the COMPARE full-gridded analysis is appropriate, we perform the following exercise. Approximately 50% of the domain consists of land and 50% is ocean, and this mask is shown in Fig. 4. We compute rms errors and bias for the same variables as for Tables 4 and 5 for E3 in each of these subdomains. Additionally, we compute these same errors as measured directly against the sounding observations. We evaluated the rms error and bias for E3 in which there are 12 participants. Each of the six parameters (Fig. 16 abscissa) is averaged for 12, 24, and 36 h for the RFE and for the 11 other participants. The percentage gain (or loss) of the RFE model against the mean of the other 11 participants is shown in the figure. The ocean mask, a region where the number of soundings is limited and in which the analysis is less certain, is also where the RFE enjoys strong performance against the other participants. However, the land mask (Fig. 16b), where the analysis is likely to be more accurate, is also a region where the RFE performs well against the other participants. The exception to this is the height bias in which the model ranks ninth and has a mean bias magnitude of 13.8 m, as opposed to the 11.6 m magnitude of the other participants. This relatively large bias in the RFE is related to positive biases, particularly at 300 hPa. The relative performance as measured against the soundings still shows a generally good performance by the RFE, with the exception of the categories of the rms error in wind speed and the

TABLE 5. Frequency of maximum/minimum scores from the soundings (parameters listed in Table 3).

| Forecast range (h)	12	12	24	24	36	36	
Experiment	1	3	1	3	1	3	
Horizontal resolution (km)	100	50	100	50	100	50	
Institution model (identifier)							Total
CSIRO LAM (C)	1/0	1/1	0/2	0/1	0/0	0/2	2/6
AES RFE (R)	0/2	1/0	1/2	0/0	0/2	0/3	2/9
AES MC2 (M)	1/0	0/0	0/0	1/1	1/2	0/0	3/3
Météo-France, Toulouse (T)	1/1	0/0	4/1	2/2	2/2	2/1	11/7
FISBAT, Italy (B)	0/2	0/2	0/3	0/3	1/1	1/1	2/12
JMA JLASM (J)	0/2	0/3	1/0	3/1	2/2	2/1	8/9
UKMO							
(E)	0/1	0/1	0/0	0/0	0/0	0/0	0/2
(K)	0/3	0/4	0/1	0/3	0/1	0/1	0/13
(A)	0/1	0/2	1/0	0/0	0/2	0/2	1/7
PSU–NCAR							
MM4 (P)	2/0	2/0	2/2	1/0	1/0	0/0	8/2
NCEP eta (N)	3/2	3/1	3/0	3/0	3/0	3/1	18/4
UW—NMS (W)	4/0	5/0	0/1	2/1	2/0	4/0	17/2
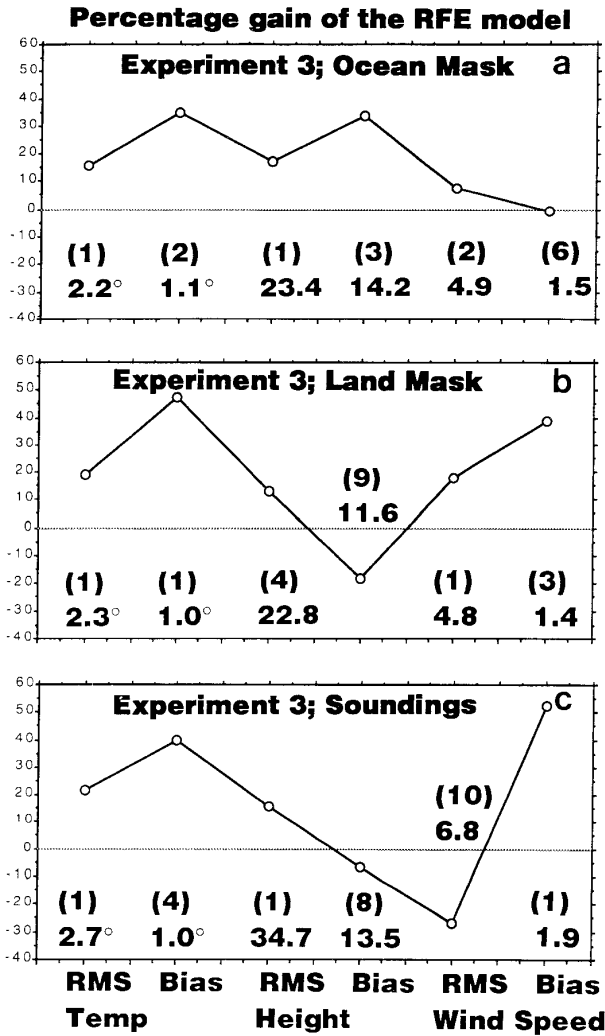
FIG. 16. Percentage gain of the RFE model in E3 vs the other 11 participants for the indicated parameters, averaged from the values at 850 and 300 hPa for 12, 24, and 36 h. The panels refer to (a) the ocean, (b) land mask, and (c) the soundings. The ranking of the RFE is indicated in parentheses. Numbers on the next row show the mean of the parameter in degrees for the temperature, m for the heights, and m s^{-1} for the wind speed rms error and biases.

height bias (again principally positive). There is a slight improvement in the performance of the participants in the more data-rich land region than over the ocean. Although the RFE model performs especially well over the data-sparse ocean, it also performs well over the land and with respect to the soundings. We conclude that the use of the gridded analysis does not markedly inflict a disadvantage onto the participants.

6. S1 scores and cyclone structure

An accepted measure of skill in forecasting gradients in geopotential height is the S1 score (Teweles and Wobus 1954). Figure 17, showing the participants'
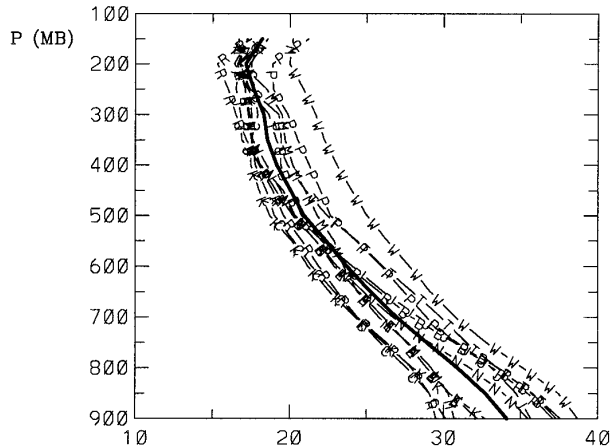

FIG. 17. Vertical profiles of S1 scores for E3 (50 km/18 levels) at 24 h for each of the 12 participants (lettered), with the composite score shown in solid.

vertical profiles at 24 h in E3, reveals a general decrease upward from a composite value of 34 at 1000 hPa to a minimum composite of 16 at 300 hPa. This upward decrease may be due to relatively large phase errors near the surface (e.g., Fig. 3). The vertical profiles of S1 composites (Fig. 18) show peak values at the 24–30-h range. One possible explanation for this early peak may be that the cyclogenesis and intensity maximize near these times and that the models have erroneously late intensification between 30 and 36 h (Fig. 2). The sensitivity of S1, composited from the four common participants, to simulation experiments (Fig. 19) at 1200 UTC 7 March shows the best results from E3, E4, E5, and E6 in the lower troposphere and the best results from 300 hPa upward to be from E6. Generally, horizontal resolutions finer than 100 km (except for the later-initialized E6) produce better results in the
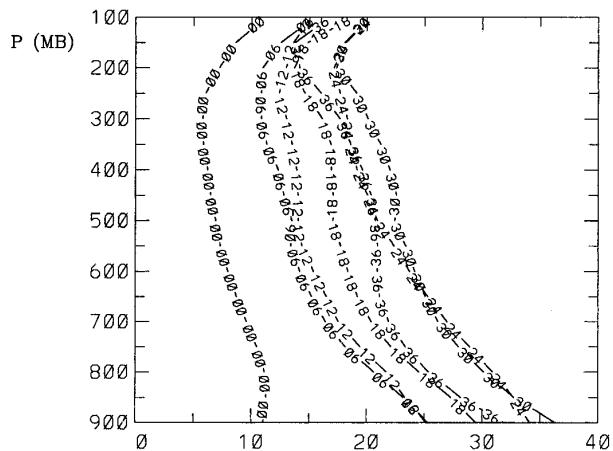

FIG. 18. Vertical profiles of composite S1 scores for E3 (each 6 h, with profile labeled in h).
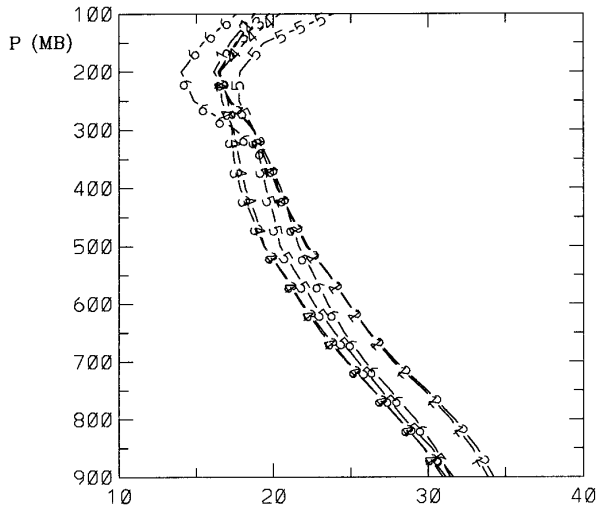
FIG. 19. Vertical profiles of S1 score, composited for each experiment (listed by number), at 1200 UTC 7 March 1986 and derived from the four models participating in all.

lower troposphere, although the differences of <4 are all small. This small difference is likely due to the fact that all model runs captured the basic cyclogenesis events, which are forced by large-scale dynamics.

The S1 scores shown in this study are generally smaller than those published earlier. Though it is likely that much variability of the scoring variability is due to the differing nature of the cases, it is interesting to consider the numbers published for earlier cases. Anthes (1983) has published a range of 45 (sea level pressure) to 20 (300 hPa). Similar scores of 44 for sea level and 24 for 500 hPa were reported by Koch et al. (1985) for NCEP's Limited Area Fine Mesh Model. More recently, Kuo et al. (1996) found S1 scores at sea level during the 1992 field program STORMFEST (Fronts Experiment and Systems Test; Cunning and Williams 1993), ranging from 33 for a 20-km version of MM4 to 38 for the 80-km eta model at NCEP. Stoss and Mullen (1995), in their study of recent NCEP NGM errors, find the 500-hPa S1 scores to range from 19 at 12 h to 31 at 36 h.

To understand the time evolution of the secondary cyclone in further detail, we compute the relative geostrophic vorticity (ζ_g), at the secondary cyclone center, for the crucial period of development from 0600 UTC 7 March through 0000 UTC 8 March in E3 and from the analysis (Fig. 2b). This vorticity is

$$\zeta_g = \nabla^2 p (f\rho)^{-1}, \qquad (4)$$

where p is the sea level pressure, f is the Coriolis parameter, and ρ is the density. We apply the horizontal Laplacian operator ∇^2 on a grid mesh centered at the cyclone and spaced four grid points (approximately 200 km) in the northward, southward, westward, and eastward directions. This 200-km distance corresponds

approximately to the scale of the system at 0600 7 March (Fig. 4), the distance from the center to the nearest col in the pressure field (Nielsen and Dole 1992). The proximity of the system to the southern boundary precludes us from computing the geostrophic relative vorticity at 0000 UTC 7 March. The results, shown in Fig. 20a, reveal a general intensification in both the analysis and in the simulations through 30 h (1800 UTC 7 March). This result is similar to what has been shown for central pressures (Fig. 2b). Though both these figures reveal the analysis intensity at 30 h to be in the middle range of the simulations, all of the
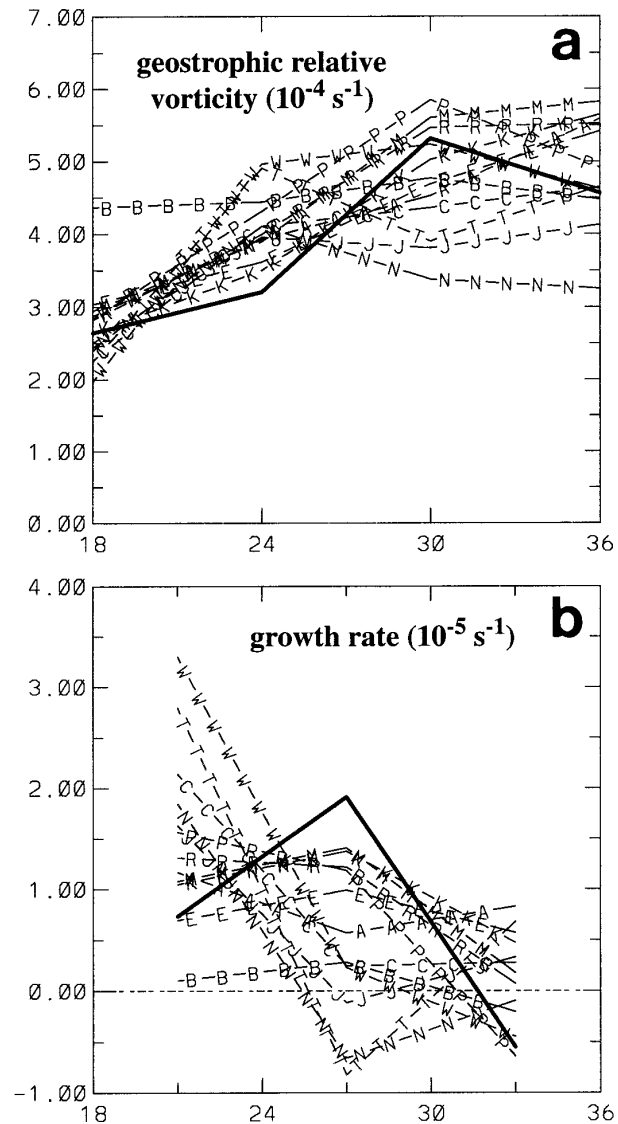

FIG. 20. (a) Time series of the geostrophic relative vorticity (10^{-4} s^{-1}) for the analysis (solid) and the participants (lettered as in previous figures) in E3 for the period from 0600 UTC 7 March (18 h) through 0000 UTC 8 March 1986 (36 h) and (b) growth rates (10^{-5} s^{-1}) for the same period and experiment.
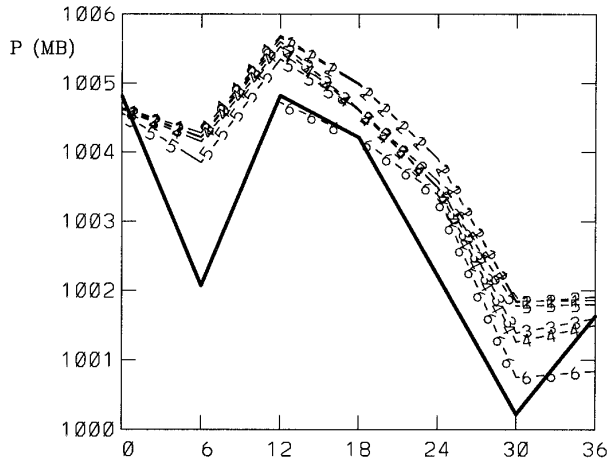
FIG. 21. Time series of domain-averaged sea level pressure (hPa) for the participants in each of the six experiments (numbered). The RPN-prepared COMPARE analysis is also shown (solid).

participants simulate a vorticity beyond that of the analysis at 24 h. There is less of a systematic bias when central pressure alone is considered (Fig. 2b). The decay from 30 to 36 h seen in the analysis and in most of the simulations (Fig. 20a) is associated with a weakening of the inner core of the pressure gradient, even though the central pressures are either constant or falling in Fig. 2b. A virtue of using the vorticity as a measure of intensity of the cyclone is that it may be used in the vorticity equation to compute a growth rate. We consider the semigeostrophic form of the vorticity equation at the cyclone center (Bluestein 1993), with the advection omitted (zero wind at the center), and with the tilting term assumed small,

$$\frac{\partial \zeta_g}{\partial t} = -(\zeta_g + f)\nabla \cdot V, \qquad (5)$$

where $\nabla \cdot V$, is the horizontal divergence. We may integrate (5) at 6-h intervals to find the growth rate

$$-\nabla \cdot V = (6 \text{ h})^{-1} \ln\left[\frac{(\zeta_g + f)_{\text{final time}}}{(\zeta_g + f)_{\text{initial time}}}\right]. \qquad (6)$$

The results, shown in Fig. 20b, reveal a wide range of growth rates among the participants, though most, including the analysis, exceed the value of 1.0 $\times 10^{-5}$ s^{-1}. This value corresponds to an e-folding time (TE) of 27.7 h, approximately 1 day, in which it takes the cyclone's intensity to amplify by a factor of 2.72. Here, TE is found from

$$\text{TE} = -(\nabla \cdot V)^{-1}. \qquad (7)$$

The e-folding time of 1 day corresponds to the time-scale of classic secondary or frontal cyclogenesis (Thorncroft and Hoskins 1990).

Figure 21 shows the domain-averaged SLP, or total mass, averaged among the four common participants

for each experiment. Generally, there is too much mass in the domain, as compared with the analysis. This systematic overestimation of mass occurs primarily over the maritime region (shown in Fig. 4), particularly after 0600 UTC 7 March, when the offshore cyclogenesis is especially active. This result is consistent with deep positive geopotential height biases found earlier (see Fig. 15c). The error is the largest among the low-resolution runs (E1 and E2). The best simulation for this parameter is E6, in which the initialization occurs 12 h later at 0000 UTC 7 March 1986. The error in sea level pressure is maximized at only 6 h into the simulations of E1–E5, likely due to more uncertainty in the analysis at this time. Nevertheless, the simulations retain their excessive mass throughout the subsequent 30 h. In contrast to the early growth of SLP error, the precipitable water (PW) errors (Fig. 22) do not grow substantively until after 12 h. There is a systematic excess of PW by nearly 1 mm at the end of the 36-h period. Surprisingly, the E6 simulations produce the largest error in the form of excess PW. This may be related to the delay in precipitation production as a result of the spinup problem (Turpeinen et al. 1990; Turpeinen 1990). This may be especially important because E6 is initialized when there is relatively large precipitation production (Fig. 26a) and the secondary cyclogenesis has begun (Fig. 2). Indeed, during the time period from 0000 through 1200 UTC 7 March, the E6 domain-averaged precipitation is less than that found in the other experiments (not shown). During the final 12-h period, the E6 precipitation exceeds that found in the other runs (not shown), so that by 0000 UTC 8 March, the E6 precipitable water is closer to the other values (Fig. 22). We speculate, therefore, that the systematic excess in PW especially toward the end of the simulation is related to a systematic underestimate in the domain-averaged precipitation.
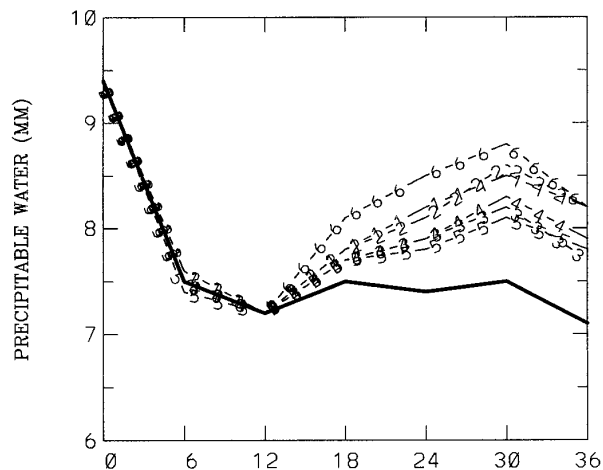

FIG. 22. Precipitable water (mm) from 1000 to 50 hPa for each of the six experiments (numbered). The RPN-prepared COMPARE analysis is also shown (solid).

7. Potential vorticity structure

The potential vorticity (PV) structure (Hoskins et al. 1985) is especially important to the early stages of the cyclogenesis. We have seen from Fig. 2c that the model simulations underestimated the central intensity of the secondary cyclogenesis at 0600 UTC 7 March 1986. Figure 23a shows the 500-hPa height and potential vorticity analysis at this crucial time. The incipient secondary low (see Fig. 4) is located about 200 km downstream of a PV maximum of 1.5 PVUs (1 PVU = 10^{-6} K kg^{-1} m^2 s^{-1}) at approximately 40°N, 71°W. The dynamic tropopause thermal structure (defined here as the 2.0 PVU surface, Fig. 23b) reveals a 331-K ridge extending through Nova Scotia that has warmed (tropopause lifting) prior to this time. This feature corresponds well with the PV minimum on the 500-hPa surface (Fig. 23a). Concurrently, the cold upstream trough, extending through New England (Figs. 23a,b) has amplified (not shown). However, the composite E3 error field for this time (Fig. 23c) exhibits a systematic damping of this wave with the temperature 6 and 8 K too warm in the respective New England and offshore cold troughs, and 18 K too cold in the warm ridge. Though individual participants produced equivalent amplitudes to the observed thermal wave at 0600 UTC 7 March, the scale and phase were at odds with the analyses. Each factor would have an important influence on this early cyclogenesis. Even with a higher vertical resolution, E4 has a comparable error in the trough and a larger error in the ridge that is 23 K too cold (not shown).

The cross section, extending through the secondary low and this upper PV maximum (Fig. 24), reveals a favorable structure for surface cyclogenesis in which the PV maximum has a coherent structure above 500 hPa, revealing a relatively low tropopause. Figure 25, showing the composite E3 cross section at the same time, reveals that the upper-level PV maxima are weaker than observed (Fig. 25b). Associated with this error is a warm bias in 2°C between and 500 and 600 hPa. To check whether this systematic deficit of upper-tropospheric PV is a consequence of simply the compositing of the participants, we examine the individual cross sections and find that the maximum 350-hPa PV values range from 1.5 to 4.0 PVUs. Clearly, the simulations from E3 all fall short of the observed PV maximum of 4.9 (Fig. 24). Though a quantitative analysis of the PV structure (in the form of an inversion) is needed to confirm the role of PV in the cyclogenesis, the available evidence suggests that the lower-level cyclogenesis simulations could have been improved with a more realistic simulation of the dynamic tropopause. This result is consistent with our earlier findings of large wind speed rms errors (Fig. 10a) and negative biases (Fig. 11a) near the tropopause.
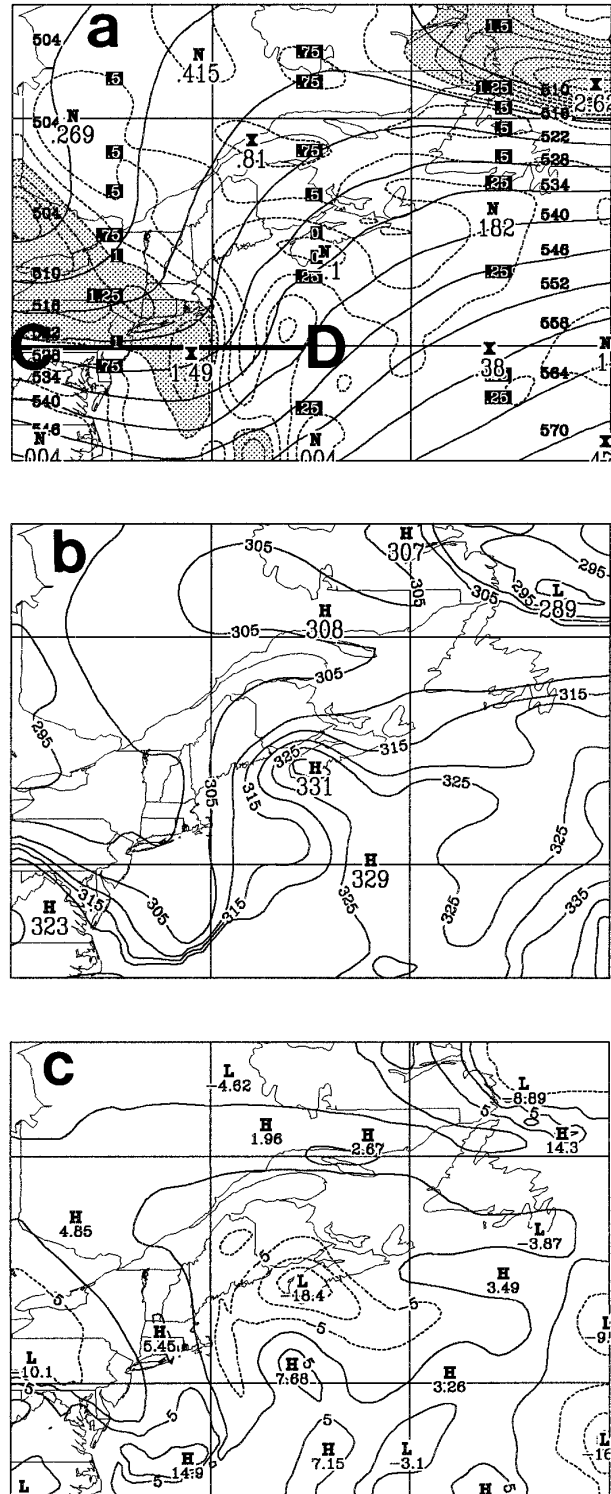

FIG. 23. (a) Height (solid, interval of 6 dam) and potential vorticity (dashed, interval of 0.25 PVUs; shaded in excess of 1.0 PVU) at 500 hPa from the analysis at 0600 UTC 7 March 1986. Cross section shown in Fig. 24 is C–D. (b) Potential temperature (interval of 5 K) on the dynamic tropopause, defined as the surface of 2.0 PVUs. (c) Composite E3 error field of potential temperature (K) on the dynamic tropopause.
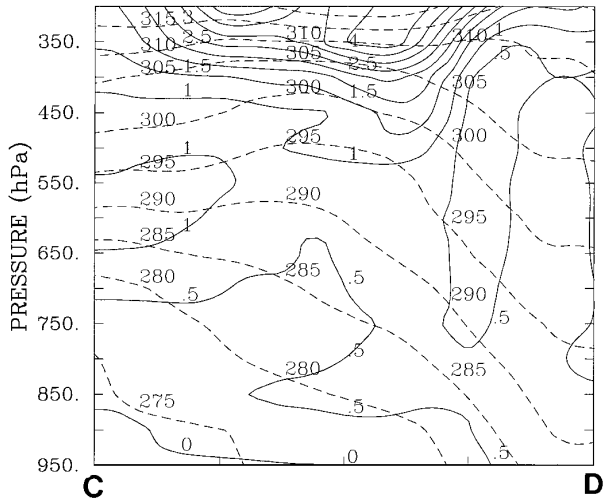
FIG. 24. Cross section (location shown in Fig. 23) of analyzed PV (dashed, interval of 0.5 PVUs) and potential temperature (solid, interval of 5 K) at 0600 UTC 7 March 1986. The PV is 38°–42°N latitudinal mean.

8. Precipitation verification

Precipitation amounts are verified in 6-h intervals for the 36-h period, ending at 0000 UTC 8 March 1986. Data for the Canadian regions consist of 6-h totals provided by Environment Canada's Climate Information Branch. The U.S. data are computed from hourly data obtained from the Data Support Section of NCAR's Scientific Computing Division. The verification domain is restricted to land stations. Figure 26 shows the domain and the analyses for the periods of 0000–0600 and 0600–1200 UTC 7 March. This example of the verifying analysis illustrates the evolution of relatively heavy precipitation along the New Brunswick and Nova Scotia coastal regions during the later period and during the secondary cyclogenesis (Figs. 2–4). A total of 475 stations are used in this study, with the most substantial number existing in the northeastern United States.

Model data are interpolated bilinearly onto the irregular grid of reporting stations. The verification statistics are calculated using categorical amounts for the 6-h periods. The categories, or threshold amounts, are 0.2, 1.25, 2.5, 6.25, and 12.5 mm. Verification scores are computed using the threat, bias, and skill scores. Further details on these and the categorical scoring may be found in Anthes (1983).

The threat score (TS) is computed as follows:

$$TS = C(F + R - C)^{-1} \times 100, \qquad (8)$$

where C is the number of stations correctly forecast to receive a threshold amount of precipitation, R is the number of stations at which the threshold amount is observed, and F is the number of stations forecasted to receive the threshold amount. Figure 27 shows the

mean threat scores for the 12 participants. The highest scores of near 40 for the measurable threshold occur in the 6–18-h range, while the larger amount thresholds peak later in the forecast. The range for the threat score for 0.2 mm is substantial, peaking at 24–30 h with values as low as 16.0 and as large as 50.4. Previously published scores of 6-h precipitation amounts include those of Corfidi and Comba (1989), in which subjective forecasts from NCEP's Meteorological Operations Division were used. The minimum amount verified in their study was 6.25 mm, so a direct comparison is not possible. However, their threat scores were in the 25–40 range for March, suggesting that the results for this
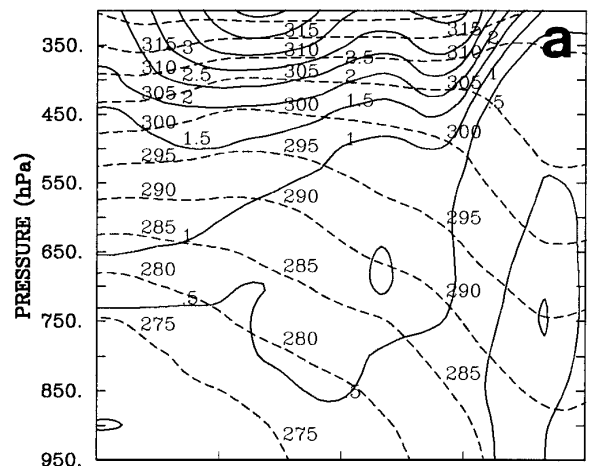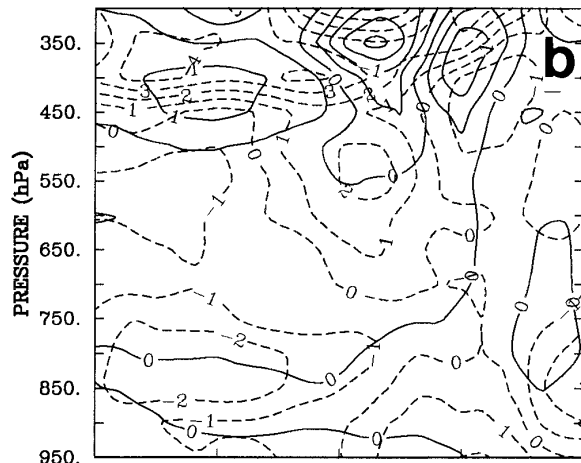

FIG. 25. Cross section (location shown in Fig. 23) of the (a) composite E3 potential temperature (dashed, interval of 5 K) and PV (solid, interval of 0.5 PVUs) and (b) composite errors in temperature (dashed, interval of 1 K) and PV (solid, interval of 0.5 PVUs) for each participant at 0600 UTC 7 March 1986.
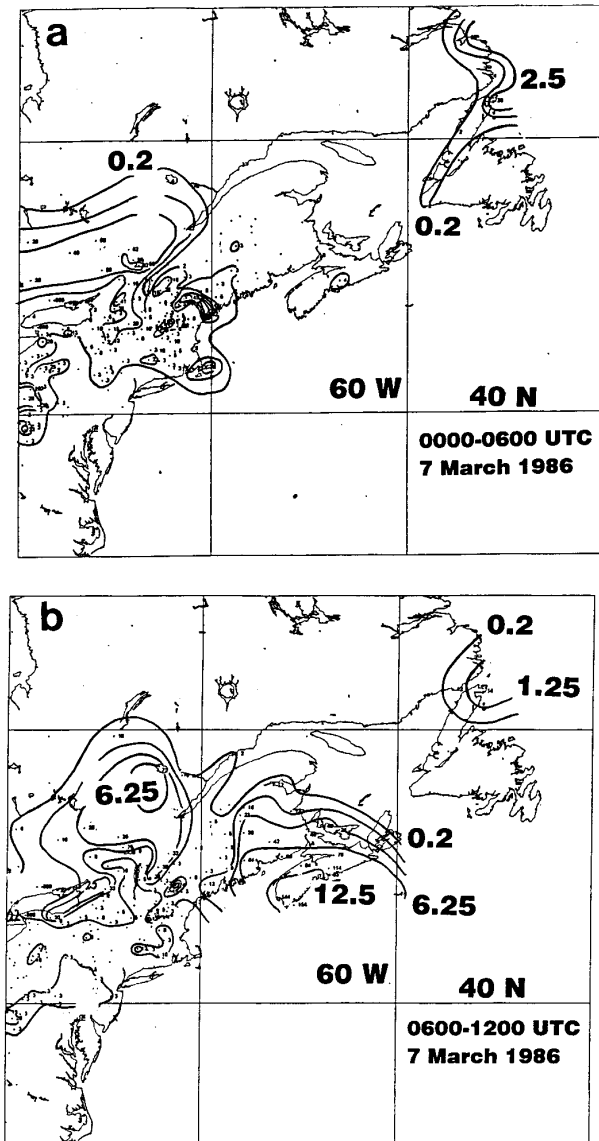
FIG. 26. Precipitation analyses (mm) with reporting stations shown as dots for (a) 0000–0600 UTC and (b) 0600–1200 UTC 7 March 1986. Contours are for 0.2, 1.25, 2.5, 6.25, and 12.5 mm.

COMPARE case are favorable, especially after 24 h of forecast range.

The skill score is defined as

$$SS = (TC - E)(T - E)^{-1}, \qquad (9)$$

where T is the total number of stations, TC is the total number of correct station forecasts (including those for less than, equal to, and beyond the threshold amount), and E is the number of correct forecasts based on some standard, such as random chance, persistence, or climatology. The skill score used here is that which uses random chance as the standard: the Heidke skill score (Heidke 1926; Brier and Allen 1951). Following Brier

and Allen (1951), we define the number E, the number of correct station forecasts expected by chance, for any category, as

$$E = [(R)(F) + (RD)(FD)]T^{-1}, \qquad (10)$$

where RD is the number of stations observing less than the threshold and FD is the number of stations forecasted to receive less than the threshold amount. One of the advantages of using the skill score is its ease of interpretation. Skill has its maximum of 1.00, which is perfection. If the skill score is zero, then the forecast makes no improvement over the random chance standard. If there is negative skill, then random chance outperforms the forecast.

Figure 28 shows the mean Heidke skill scores for E3. While skill for the threshold 0.2 mm shows a decline with increasing range, the scores for the next two higher amounts indicate the best skill is attained after 18 h of the model integration. As is the case for the threat score, the range in the skill score is considerable, with extreme values of .00 and .59 at 24–30 h for the 0.2-mm threshold (not shown).

The bias score, defined as

$$B = \frac{F}{R}, \qquad (11)$$

where B of one indicates no bias, and values exceeding one indicates the model overforecasts the frequency of the threshold amount, and a bias less than one means the particular amount is underforecast. The bias results, shown in Fig. 29, show a tendency for the models to overforecast the measurable threshold; this tendency increases with increasing forecast range. Larger amounts are also overforecast throughout the forecast period. The range is especially large at 24–30 h for 0.2 mm, with extreme values of 1.36 and 6.05 (not shown). After 6 h, all of the participants showed biases exceeding 1.0.

Figure 30 shows the mean Heidke skill scores for the 10 common participants in E1, E2, E3, and E4, according to forecast range. The greatest improvement in skill for the 0.2-mm threshold occurs as a consequence of increasing the horizontal resolution from 100 to 50 km. However, this improvement does not occur until after 12 h into the integrations. Negligible skill improvement, compared with the other simulations, is observed when considering the highest-resolution forecast, E5 (not shown).

To understand the relationship, if any, between performance in simulating the mass and precipitation fields, we compare the time-averaged (for 12, 24, and 36 h) rms errors in E3 at 850 and 300 hPa for each participant in E3 with the time-averaged (for the 0–6-, 6–12-, ... 30–36-h periods) Heidke skill scores in predicting a threshold amount of 0.2 mm in E3. The results, shown in Fig. 31, *show little relationship between strong performance in predicting the mass field*
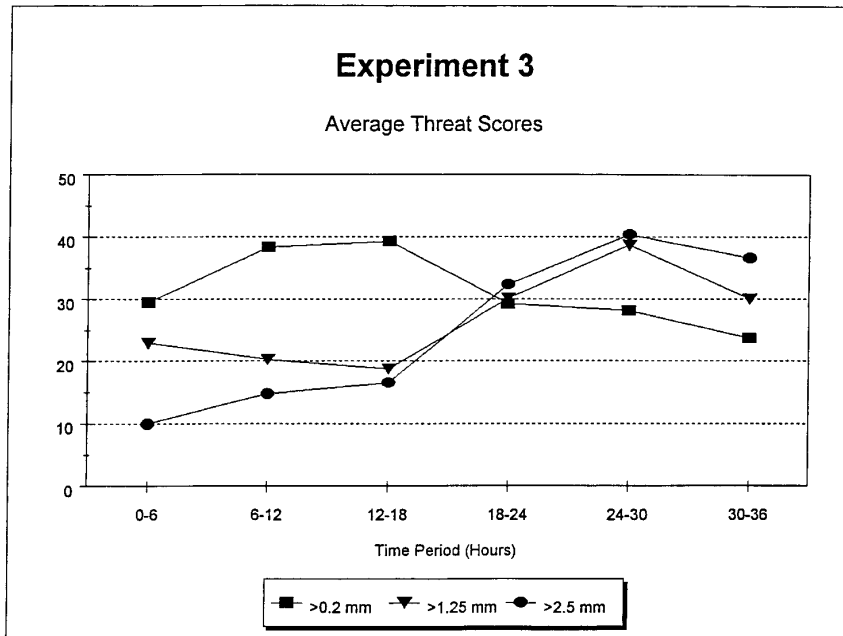
Experiment 3

Average Threat Scores

FIG. 27. E3 threat scores, averaged for the 12 participants for the 0.2-, 1.25-, and 2.5-mm threshold amounts, as a function of forecast period.

with that performance in skillfully predicting a measurable amount of precipitation. Several of the participants with relatively exceptional rms errors have near-zero skill in precipitation prediction. Conversely, participants with exceptionally high Heidke skill scores do not always show low rms errors in the height field, particularly at 850 hPa. The linear regressions, also shown, confirm our conclusion. The correlation coef-

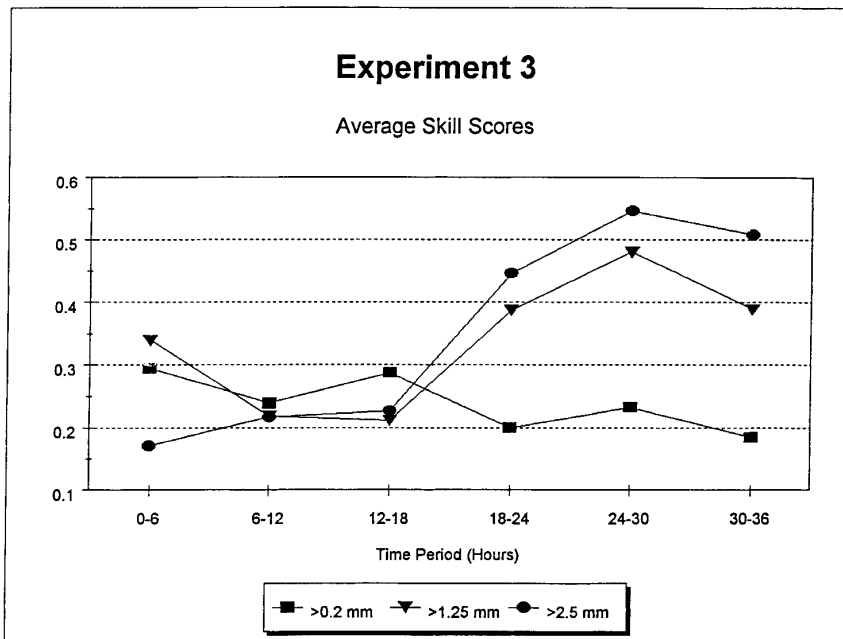## Experiment 3

Average Skill Scores

FIG. 28. E3 Heidke skill scores, averaged for the 12 participants at the 0.2-, 1.25-, and 2.5-mm threshold amounts, as a function of forecast period.
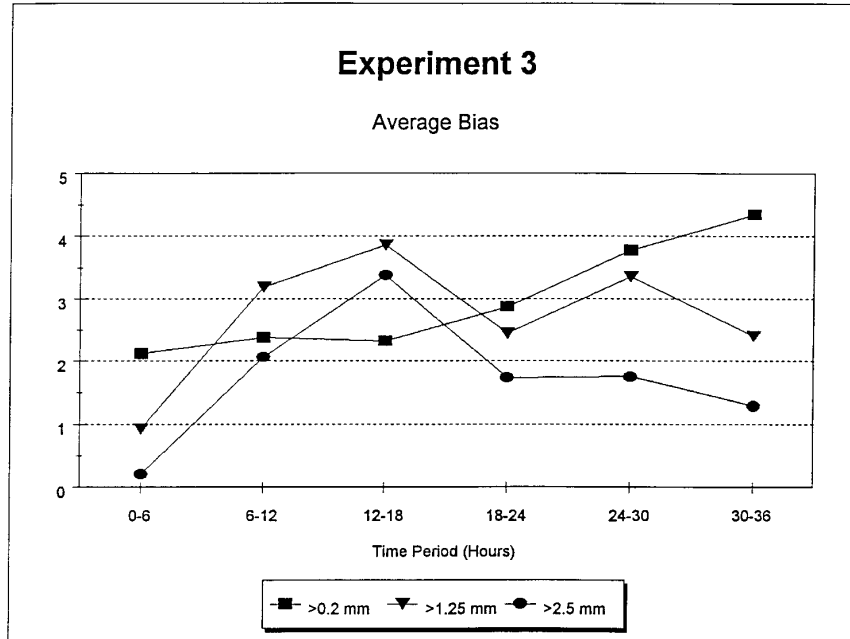
FIG. 29. E3 bias scores, averaged for the 12 participants at the 0.2-, 1.25-, and 2.5-mm threshold amounts, as a function of forecast period.

ficient of .44 between the 300-hPa rms error and the Heidke skill score (Fig. 31b) is not statistically significant. This result is consistent with the recent conclusion of Roebber and Bosart (1994) in which they show a substantial temporal increase in 250-hPa rms error in the vector wind from 1982 through 1991 from Kalnay
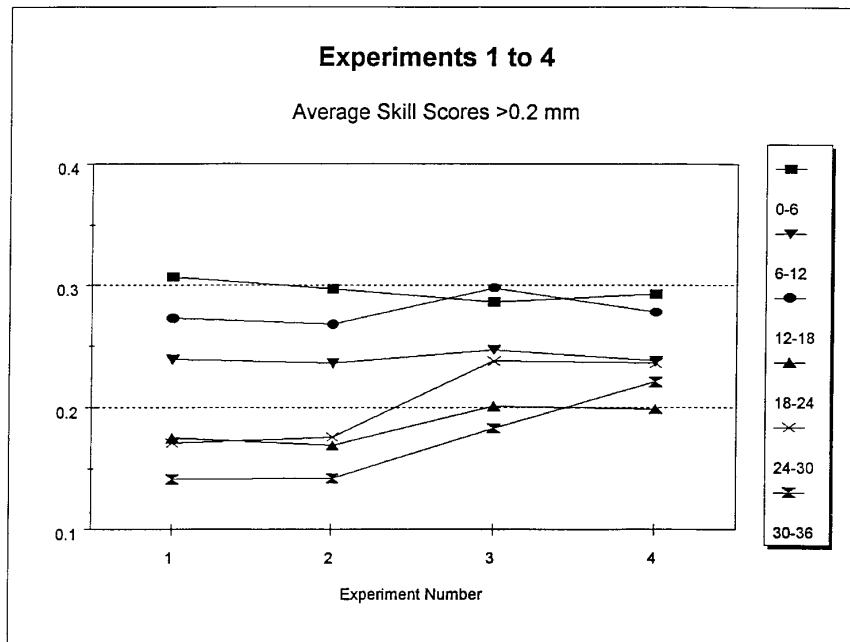

FIG. 30. Mean Heidke skill score for the 0.2-mm threshold, averaged among the 10 common participants for E1–E4, according to the forecast range (0–6, 6–12, 12–18, 18–24, 24–30, and 30–36 h).
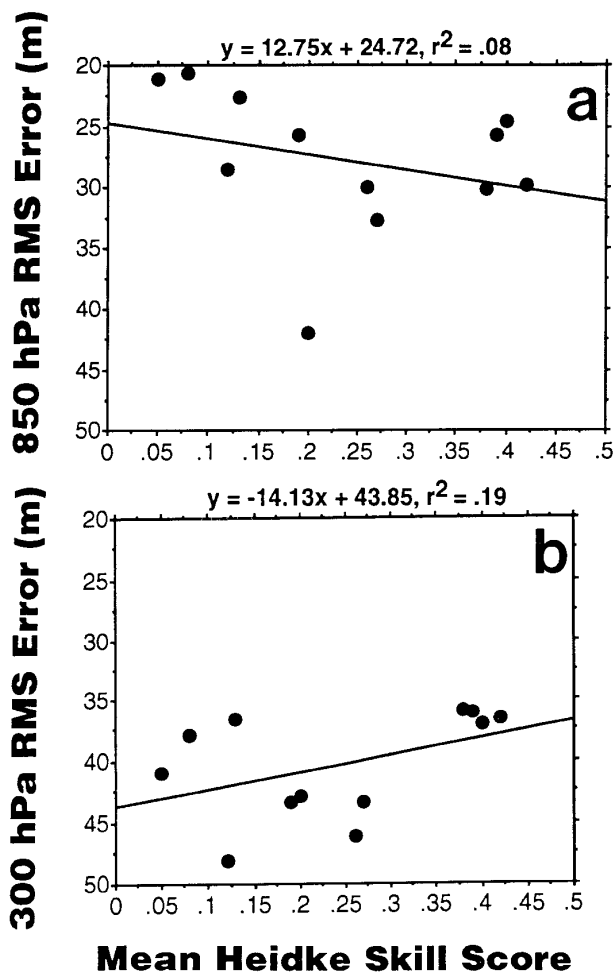
FIG. 31. Scattergrams of time-averaged Heidke skill score for a threshold of 0.2 mm (abscissa) vs (a) time-averaged 850-hPa rms error (m, ordinate), and (b) 300-hPa rms error (m, ordinate). Linear regression line and equation are also shown on each panel, with correlation coefficient squared (r^2) also shown.

et al. (1990) and little discernible trend in precipitation probability forecasting skill during the same period (Bosart 1983). *Clearly, a good performance in predicting the mass and wind fields does not imply similar results in precipitation forecasting.*

9. Conclusions

We have presented the results of a model intercomparison project designed for a case of rapidly intensifying secondary cyclogenesis. We find that even at the lowest resolution (100 km/18 levels), a secondary cyclogenesis is captured by all of the participants (Fig. 2a). *The most substantial improvement in central pressure forecasts occurs as a consequence of increasing the horizontal resolution from 100 to 50 km, with negligible improvement from increasing the vertical resolution (Fig. 2c).* This result is consistent with earlier

research showing that horizontal resolution increases have more impact than vertical resolution increases (Kuo and Low-Nam 1990). Though Lindzen and Fox-Rabinovitz (1989) have suggested that vertical resolution in models should be increased, the surface cyclogenesis forecasts are relatively insensitive to such changes, at least within the context of the COMPARE experimental design.

Especially strong variability, among the participants, in rms wind speed, height, and temperature errors occur near the locations of the upper- and lower-level jets (Fig. 10), with a systematic negative bias in wind speed (Fig. 11a) and temperature (Fig. 11b), and a positive bias in heights (especially at lower levels, Fig. 11c). Such biases generally persist after 12 h into the integrations (Fig. 13). *The largest improvement in model performance (defined in terms of central pressures, rms errors, and biases of wind speed, height, and temperature) occurs as a result of increasing the horizontal resolution from 100 to 50 km (Figs. 14 and 15). The use of the very highest resolution models for this case (E5) produced little improvement in performance in these parameters, compared with the 50-km simulations.* This does not imply that increasing the horizontal resolution cannot have an important impact on cyclone intensification. We believe that this result would be determined by improved model physics, especially at the cloud scale. We find that S1 scores improve with increasing horizontal resolution, even to the highest resolution (E5), with particularly more improvement above 300 hPa at the later initialization time (E6).

We find substantial variability in placement of the secondary cyclone (Fig. 3), with a suggestion of systematic westward displacement that has also been found by Oravec and Grumm (1993) for the NCEP model during the 1991 cold season. Our finding of the excessively weak/strong offshore/inshore surface low generalizes to excessively high/low surface pressures offshore/inshore (Fig. 6) and to the excessively cold/warm tropospheric temperatures offshore/inshore (Fig. 7). Additionally, the continental composite anticyclone is forecast to be too weak by 2 hPa (Fig. 8) and too far by 300–600 km to the southeast (Fig. 4). One of the consequences of the excessive warmth of this anticyclone, and the excessive coldness of the marine regions (Fig. 7), is to weaken the background baroclinity, relative to the observations (Fig. 9). Furthermore, the systematically weak tropopause thermal perturbation (Figs. 22–24) may have contributed to the marine atmosphere's weak cyclogenesis.

The precipitation bias scores indicate a tendency for all models to overforecast the smaller verifying precipitation amounts. *Results for the measurable threshold (0.2 mm) indicate the largest gain in scores results from increasing the horizontal resolution from 100 to 50 km. Negligible benefit occurs as a consequence of increasing the resolution to 25 km.* These precipitation

results are consistent with our mass field scores, in which the most obvious resolution benefit results from increasing the horizontal resolution from 100 km in E1 to 50 km in E3. An interesting result of this study is that *there is little correlation between exemplary performance in forecasting the mass field and equivalent achievement in precipitation forecasting (Fig. 31)*.

The scientific issues, elucidated in section 2, have been partially addressed by this study. We have demonstrated that there is a systematically weak thermal wave on the dynamic tropopause at the crucial early stages of the cyclogenesis. Its associated 500-hPa cyclonic vorticity maxima have been found by Sanders (1986, 1988) to be a crucial precursor to explosive cyclogenesis along the east coast of North America. This error may be related to the weaker-than-observed surface system at this time. We have shown the planetary boundary layer to be too statically stable, too cold, and its baroclinity to be too weak. All of these errors would suppress ascent, precipitation, and frontogenesis. Our finding that the offshore regions are generally too cold and too stable suggests that sensible heat fluxes may be too weak. However, this speculation, even if verified, may or may not be relevant to the secondary cyclogenesis. The low-level jet amplitudes are highly variable (not shown), and the participants with the strongest lower-level winds did not always produce the best cyclogenesis forecast. The successful simulation of the system may be best related to the proper phasing of the upper and lower cyclonic disturbances.

Our study has produced new scientific issues relating to the secondary cyclogenesis. First, a component of the weak baroclinicity is related to the participants' tendency to produce an excessively weak and warm anticyclone to the north. Second, the factors that produce the upstream upper-level PV maximum need to be clarified. In particular, the models' simulations of upper waves with varying amplitudes suggests that physics (e.g., diffusion, turbulent mixing, or radiation) or the varying treatments of the upper-boundary conditions, could be playing a role. We can only speculate that an improved simulation of the upper PV structure and associated cyclogenesis may require especially high resolution afforded by isentropic coordinates in the vicinity of the dynamic tropopause. Third, we find a substantive systematic overdeepening of the inshore primary cyclone. This finding suggests that the cyclolysis mechanism(s) is/are not being properly represented in the models.

An additional related issue arising from this study is the role of initial analysis in the performance of the models. The RFE model, when scored against its own verifying analysis, performed exceptionally well (Table 4, Figs. 16a,b). However, when the models were scored in a representative range of accepted parameters against the sounding observations, the model's performance is weaker (Table 5); however, this weakness is reflected only in the height bias and wind speed rms

error (Fig. 16c). As indicated in Fig. 16, the accuracy of the verifying analysis over the oceans also needs to be considered in the verification. The model generating the analysis will tend to have less errors than other models in regions that lack a dense upper-air network, such as over the ocean. Future model intercomparisons should attempt to have a dense observational network over the whole verification domain. The sensitivity of these results to the analysis suggests that the initial conditions may be playing a crucial role in determining the model simulation of this mesoscale cyclogenesis case. Future work is being directed toward understanding this role.

REFERENCES

Anthes, R. A., 1983: Regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.,* **111,** 1306–1335.
——, E.-Y. Hsie, and Y.-H. Kuo, 1987: Description of the Penn State/NCAR Mesoscale Model Version 4 (MM4). NCAR Tech. Note NCAR/TN-282+STR, 66 pp. [Available from NCAR Publications Office, P.O. Box 3000, Boulder, CO 80307-3000.]

Arakawa, A., 1972: Parameterization of cumulus convection. Design of the UCLA general circulation model. Numerical simulation of weather and climate. Tech. Rep. 7. [Available from Dept. of Meteorology, University of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90024-1565.]

Benoit, R., J. Côté, and J. Mailhot, 1989: Inclusion of a TKE boundary layer parameterization in the Canadian Regional Finite Element Model. *Mon. Wea. Rev.,* **117,** 1726–1750.
——, S. Pellerin, and W. Yu, 1996: MC2 Model performance during the Beaufort and Arctic Storm Experiment. *Atmos.–Ocean,* in press.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112,** 693–709.

Bluestein, H. B., 1993: *Observations and Theory of Weather Systems.* Vol. 2, *Synoptic-Dynamic Meteorology in Midlatitudes.* Oxford University Press, 594 pp.

Bosart, L. F., 1983: An update on trends in skill of daily forecasts of temperature and precipitation at the State University of New York at Albany. *Bull. Amer. Meteor. Soc.,* **64,** 346–354.

Bougeault, P., and P. LaCarrère, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.,* **117,** 1872–1890.

Brier, G. W., and R. A. Allen, 1951: Verification of weather forecasts. *Compendium of Meteorology,* T. Malone, Ed., Amer. Meteor. Soc., 841–848.

Buzzi, A., M. Fantini, P. Malguzzi, and F. Nerozzi, 1994: Validation of a limited-area model in cases of Mediterranean cyclogenesis: Surface fields and precipitation scores. *Meteor. Atmos. Phys.,* **53,** 137–153.

Chouinard, C., J. Mailhot, H. L. Mitchell, A. Staniforth, and R. Hogue, 1994: The Canadian regional data assimilation system: Operational and research applications. *Mon. Wea. Rev.,* **122,** 1306–1325.

Corfidi, S. F., and K. E. Comba, 1989: The meteorological operations division of the National Meteorological Center. *Wea. Forecasting,* **4,** 343–366.

Cullen, M. J. P., and T. Davies, 1991: A conservative split-explicit integration scheme with fourth-order horizontal advection. *Quart. J. Roy. Meteor. Soc.,* **117,** 993–1002.

Cunning, J. B., and S. F. Williams, 1993: U.S. Weather Research Program STORM-FEST Operations Summary and Data Inventory, 389 pp. [Available from U.S. Weather Research Program Office and UCAR Office of Field Project Support, NCAR, P.O. Box 3000, Boulder, CO 80307-3000.]

Dirks, R. A., J. P. Kuettner, and J. A. Moore, 1988: Genesis of Atlantic Lows Experiment (GALE): An overview. *Bull. Amer. Meteor. Soc.,* **69,** 148–160.

Emanuel, K. A., 1991: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.,* **48,** 2313–2335.

Gadd, A. J., and J. F. Kears, 1970: Surface exchanges of sensible and latent heat in a 10-level model atmosphere. *Quart. J. Roy. Meteor. Soc.,* **96,** 297–308.

Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.,* **118,** 1483–1506.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterization schemes. *Mon. Wea. Rev.,* **121,** 764–787.

Grumm, R., and J. R. Gyakum, 1986: Systematic surface anticyclone errors in NMC's limited area fine mesh and spectral models during the winter of 1981–82. *Mon. Wea. Rev.,* **114,** 2329–2343.

Heidke, P., 1926: Berechnung des Erfolges und der Gütte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geog. Ann. Stockh.,* **8,** 310–349.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.,* **111,** 877–946.

Imbard, M., and Coauthors, 1987: The PERIDOT fine-mesh numerical weather prediction system: Description, evaluation and experiments. *Short- and Medium-Range Numerical Weather Prediction,* T. Matsuno, Ed., Meteor. Soc. Japan, 455–565.

Junker, N. W., J. E. Hoke, and R. H. Grumm, 1989: Performance of NMC's regional models. *Wea. Forecasting,* **4,** 368–390.

Kalnay, E., M. Kanamitsu, and W. E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Amer. Meteor. Soc.,* **71,** 1410–1428.

Koch, S. E., W. C. Skillman, P. J. Kocin, P. J. Wetzel, K. F. Brill, D. A. Keyser, and M. C. McCumber, 1985: Synoptic scale forecast skill and systematic errors in the MASS 2.0 model. *Mon. Wea. Rev.,* **113,** 1714–1737.

Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. Atmos. Sci.,* **31,** 1232–1240.

Kuo, Y.-H., and S. Low-Nam, 1990: Prediction of nine explosive cyclones over the western Atlantic Ocean with a regional model. *Mon. Wea. Rev.,* **118,** 3–25.

——, J. R. Gyakum, and Z. Guo, 1995: A case of rapid continental mesoscale cyclogenesis. Part I: Model sensitivity experiments. *Mon. Wea. Rev.,* **123,** 970–997.

Leary, C., 1971: Systematic errors in operational National Meteorological Center primitive-equation surface prognoses. *Mon. Wea. Rev.,* **99,** 409–413.

Lindzen, R. S., and M. Fox-Rabinovitz, 1989: Consistent vertical and horizontal resolution. *Mon. Wea. Rev.,* **117,** 2475–2583.

Mailhot, J., and C. Chouinard, 1989: Numerical forecasts of explosive winter storms: Sensitivity experiments with a meso-α scale model. *Mon. Wea. Rev.,* **117,** 1311–1343.

——, and Coauthors, 1995: Changes to the Canadian Regional Forecast System: Description and evaluation of the 50-km version. *Atmos.–Ocean,* **33,** 55–80.

McGregor, J. L., 1993: Economical determination of departure points from semi-Lagrangian models. *Mon. Wea. Rev.,* **121,** 221–230.

Mellor, G. L., and T. Yamada, 1974: A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.,* **31,** 1791–1806.

Mesinger, F., Z. I. Janjic, S. Nickovic, D. Gavrilov, and D. G. Deaven, 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev.,* **116,** 1493–1518.

Nielsen, J. W., and R. M. Dole, 1992: A survey of extratropical cyclone characteristics during GALE. *Mon. Wea. Rev.,* **120,** 1156–1167.

Oravec, R. J., and R. H. Grumm, 1993: The prediction of rapidly deepening cyclones by NMC's Nested Grid Model: Winter 1989–Autumn 1991. *Wea. Forecasting,* **8,** 248–270.

Roebber, P. J., and L. F. Bosart, 1994: How sensitive is the precipitation associated with baroclinic winter storms to small variations in the synoptic-scale circulation? An analysis of observed data using regional analogues. *Proc. Int. Symp. on the Life Cycles of Extratropical Cyclones,* Bergen, Norway, Norwegian Geophysical Society, 215–220.

Sanders, F., 1986: Explosive cyclogenesis in the west-central North Atlantic Ocean, 1981–84. Part I: Composite structure and mean behavior. *Mon. Wea. Rev.,* **114,** 1781–1794.

——, 1988: Life history of mobile troughs in the upper westerlies. *Mon. Wea. Rev.,* **116,** 2629–2648.

——, and J. R. Gyakum, 1980: Synoptic-dynamic climatology of the ''bomb.'' *Mon. Wea. Rev.,* **108,** 1589–1606.

Segami, A., K. Kurihara, H. Nakamura, M. Ueno, I. Takano, and Y. Tatsumi, 1989: Operational mesoscale weather prediction with Japan Spectral Model. *J. Meteor. Soc. Japan,* **67,** 907–924.

Silberberg, S. R., and L. F. Bosart, 1982: An analysis of systematic cyclone errors in the NMC LFM-II model during the 1978–79 cool season. *Mon. Wea. Rev.,* **110,** 254–271.

Stewart, R. E., and N. R. Donaldson, 1989: On the nature of rapidly deepening Canadian East Coast winter storms. *Atmos.–Ocean,* **27,** 87–107.

——, R. W. Shaw, and G. A. Isaac, 1987: Canadian Atlantic Storms Program: The meteorological field project. *Bull. Amer. Meteor. Soc.,* **68,** 338–345.

Stoss, L. A., and S. L. Mullen, 1995: The dependence of short-range 500-mb height forecasts on the initial flow regime. *Wea. Forecasting,* **10,** 353–368.

Tanguay, M., A. Simard, and A. Staniforth, 1989: A three-dimensional semi-Lagrangian scheme for the Canadian regional finite element forecast model. *Mon. Wea. Rev.,* **117,** 1861–1871.

——, A. Robert, and R. Laprise, 1990: A semi-implicit semi-Lagrangian fully compressible regional forecast model. *Mon. Wea. Rev.,* **118,** 1970–1980.

Teweles, S., Jr., and H. B. Wobus, 1954: Verification of prognostic charts. *Bull. Amer. Meteor. Soc.,* **35,** 455–463.

Thorncroft, C. D., and B. J. Hoskins, 1990: Frontal cyclogenesis. *J. Atmos. Sci.,* **47,** 2317–2336.

Tripoli, G. J., 1992: A nonhydrostatic mesoscale model designed to simulate scale interaction. *Mon. Wea. Rev.,* **120,** 1342–1359.

Turpeinen, O. M., 1990: Diabatic initialization of the Canadian Regional Finite-Element (RFE) model using satellite data. Part II: Sensitivity to humidity enhancement, latent-heating profile and rain rates. *Mon. Wea. Rev.,* **118,** 1396–1407.

——, L. Garand, R. Benoit, and M. Roch, 1990: Diabatic initialization of the Canadian Regional Finite-Element (RFE) model using satellite data. Part I: Methodology and application to a winter storm. *Mon. Wea. Rev.,* **118,** 1381–1395.

Wood, N., and P. J. Mason, 1993: The pressure force induced by neutral turbulent flow over hills. *Quart. J. Roy. Meteor. Soc.,* **119,** 1233–1268.

Yau, M. K., and M. Jean, 1989: Synoptic aspects and physical processes in the rapidly-intensifying cyclone of 6–8 March 1986. *Atmos.–Ocean,* **27,** 59–86.

Zhang, D.-L., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer—Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.,* **21,** 1594–1609.