Evaluation of Real-Time High-Resolution MM5 Predictions over the Great Lakes Region

SHIYUAN ZHONG AND HEE-JIN IN

Department of Geosciences, University of Houston, Houston, Texas

XINDI BIAN, JOSEPH CHARNEY, WARREN HEILMAN, AND BRIAN POTTER

North Central Research Station, USDA Forest Service, East Lansing, Michigan

(Manuscript received 23 March 2004, in final form 10 August 2004)

ABSTRACT

Real-time high-resolution mesoscale predictions using the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5) over the Great Lakes region are evaluated for the 2002/03 winter and 2003 summer seasons using surface and upper-air observations, with a focus on near-surface and boundary layer properties that are important for applications such as air quality and fire weather predictions.

The summer season predictions produce a cold bias in maximum daily temperature and a warm bias in minimum temperature that together lead to a good prediction of daily mean temperature but a smaller-than-observed diurnal temperature cycle. In winter, the predicted near-surface temperatures are lower both day and night, yielding good agreement with the observed amplitude of the diurnal temperature cycle but relatively large cold bias in daily mean temperature. The predicted temperatures in the boundary layer are also systematically lower than the observed temperatures in the two seasons. The cold bias is consistent with the wetter-than-observed lower atmosphere in the model prediction, which in turn can be attributed to an inadequate specification of soil moisture. In both seasons, the model produced substantially more precipitation in all categories, especially in the heavy precipitation category, and the overprediction is primarily associated with more widespread area coverage in the model prediction. The chances of producing a false precipitation forecast are substantially higher than missing an observed precipitation event. Small systematic errors are found in the predictions of low-level winds, but above the boundary layer, the predicted winds are predominantly from the west, while the observed winds are from the west-northwest. The model is able to capture the general development and evolution of the lake–land breezes in areas surrounding Lake Michigan during summer, although errors exist in the strengths of the breezes and the timing of their transition.

Predicted early morning inversions are slightly stronger than observed in winter and weaker than observed in summer. The weak summer morning inversion results in a rapid inversion breakup followed by an earlier growth of a mixed layer after sunrise. Despite the head start, the predicted mixed-layer heights in late afternoon are lower than those observed, suggesting that either the predicted surface sensible heat flux may be too low or the boundary layer flux divergence may be too high.

Decreasing horizontal grid spacing from 12 to 4 km results in little improvement in the predictions of near-surface and boundary layer properties except for precipitation, for which the model bias is significantly reduced by the increase in horizontal resolution. The cold and wet biases and errors in inversion strengths and mixed-layer development call for extra caution when using products from mesoscale forecasts in applications such as air pollution and fire weather prediction.

1. Introduction

Mesoscale models, such as the Regional Atmospheric Modeling System (RAMS; Pielke et al. 1992) and the Pennsylvania State University–National Center for Atmospheric Research Fifth-Generation Mesoscale Model (MM5; Grell et al. 1994) are playing an increasingly important role in operational numerical weather forecasting throughout the world, thanks to the proliferation of inexpensive high-performance computers, massively parallel architecture, and distributed-memory codes. The use of high-resolution mesoscale models can add value to the operational weather forecast process, particularly in areas where topography and land-use heterogeneity modulate synoptic weather to produce localized weather phenomena. Validating

Corresponding author address: Dr. Sharon Zhong, Department of Geosciences, University of Houston, 312 S&R Building 1, 4800 Calhoun Rd., Houston, TX 77204-5007.
E-mail: szhong@uh.edu

the skill of these mesoscale models, which is important to gauge improvements in model performance, has received increasing attention in recent years.

A number of studies have evaluated the performance of mesoscale forecasting using either routine meteorological observations or data from short-term field campaigns. Many of these studies have focused on quantitative precipitation forecasting (QPF) because high-resolution mesoscale models offer great potential for improved QPF, which has long been the most challenging problem in numerical weather prediction. Chien et al. (2002), Colle et al. (1999, 2000, 2003a,b), Chen et al. (2002), and Aves et al. (2002), among others, evaluated QPF from several operational mesoscale models using observed precipitation and various statistical techniques. Other studies focused the forecast validation on a particular weather event or phenomenon such as flash flooding and winter storms (Wang et al. 2002; Gerard and Listemaa 2002; Bukovsky et al. 2002).

Most validation studies of mesoscale model predictions are limited either to a few selected cases or to a relatively short time period when special observations from field studies were available (Manning and Davis 1997; Zhong and Fast 2003). Only a few studies have evaluated mesoscale model forecasts using relatively long-term observations from several seasons and over multiple years. Using surface observations from over a 2-yr period, Colle et al. (1999, 2000) evaluated the performance of real-time MM5 forecasts for the Pacific Northwest with a focus on precipitation. They found that the forecasting skill was improved when model grid spacing was decreased from 36 to 12 km, but a further decrease of horizontal grid spacing from 12 to 4 km resulted in little improvement in forecast accuracy, although 4-km grid spacing produced a more realistic distribution of precipitation. Similar findings were obtained in Colle et al. (2003a,b) when MM5 forecasts were validated against observations over a period of nearly 2 yr for the northeastern United States.

As part of an initiative by the U.S. Department of Agriculture (USDA) Forest Service to expand existing knowledge of fire–atmosphere interactions and enhance the ability to predict and respond to the danger of wildfires, MM5 has been run twice daily in real time since the beginning of summer 2002 at the USDA Forest Service's Eastern Area Modeling Consortium (EAMC) in East Lansing, Michigan. Hourly mesoscale prediction results at high resolution for the Great Lakes region and for New England have been archived for validation purposes. This paper presents validations of MM5 predictions for the 2002/03 winter season (December–February) and 2003 summer season (June–August). The results were validated against both surface and upper-air observations in the states neighboring the Great Lakes, and the verifications were performed not only for variables that are traditionally included in forecasting validations, such as precipitation and surface temperature, but for properties that are important for air pollution and fire weather, such as mixed-layer heights, inversion strengths, moisture content in the lower atmosphere, and lake–land breezes. The hourly archive of model output allows for adequate verification of model predictions of the diurnal variations, in addition to seasonal cycles. The current study not only provides performance evaluation of high-resolution MM5 predictions on a relatively long-term basis for the Great Lakes region for the first time, but also contributes to our understanding of the strengths and weakness of mesoscale prediction. Because of the significant differences in topography, land use and type, and synoptic regimes between the Great Lakes region of the current study and the Pacific Northwest and the Northeast, where performance evaluations of high-resolution MM5 forecasting have been carried out, the results from this study bring a new perspective to the open discussion regarding the values of improved resolution in real-time numerical weather forecasting (Mass and Kuo 1998).

2. Model physics and grid configuration

MM5 is a nonhydrostatic, primitive equation mesoscale model that employs a terrain-following pressure coordinate system. MM5 can be run with two-way nested grids and with data assimilation. The model includes various physical parameterizations for radiative transfer, cloud microphysics, cumulus convection, boundary layer turbulence, and land surface exchange processes. For the EAMC's real-time prediction, MM5 is configured with four domains. The outer domain has a horizontal grid spacing of 36 km and covers the continental United States and the adjacent coastal waters as well as part of southern Canada. The middle domain, with 12-km grid spacing, encompasses the north-central/northeastern United States. Nested within the 12-km resolution domain are two domains with 4-km grid spacing: the eastern 4-km domain covers the New England area, while the western 4-km domain encompasses the Great Lakes and the neighboring states. The locations of the four domains are shown in Fig. 1. The nesting between the 36- and 12-km grids is two-way interactive, which means that solutions from the inner grid (12-km grid) feed back to its parent grid (36 km grid), but the nesting between the 12- and 4-km grids is one way, where results from the 4-km grids do not affect those in the 12-km grid. In the vertical, 35 unevenly spaced sigma levels are employed with vertical grid spacing stretched from approximately 10 m above surface to 1500 m at the model top near 12 km.

The model physics employed for the operational predictions include the cloud radiation scheme (Dudhia 1989), the mixed-phase cloud microphysics (Reisner et al. 1998), the Kain–Fritsch cumulus parameterization (Kain and Fritsch 1990), the Eta Model boundary layer parameterization (Janjic 1990), and a simple multilayer soil model.

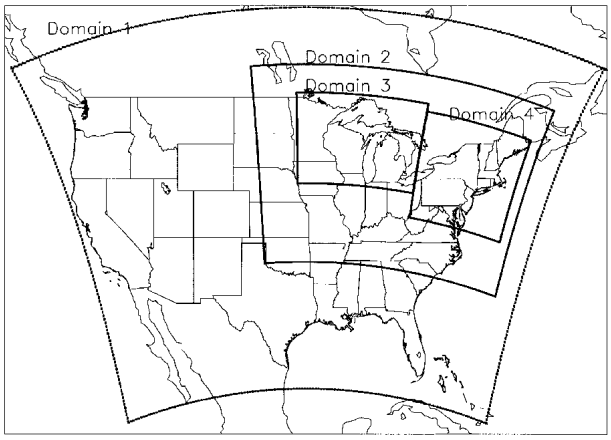

FIG. 1. The MM5 real-time forecasting domains.

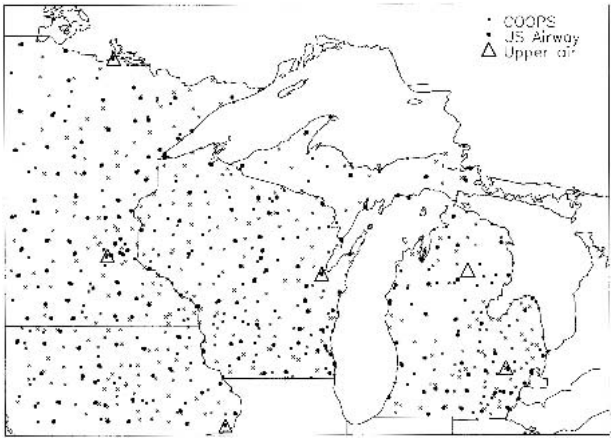

FIG. 2. Surface and upper-air stations used for the verification of MM5 forecasts on the western 4-km grid over the Great Lakes region. Symbols of the observational data types are provided in the box in the upper-right corner.

MM5 is run twice daily, initialized at 0000 and 1200 UTC using the National Centers for Environmental Prediction's (NCEP's) operational Eta Model output, on a 32-processor Linux PC cluster, which produces 48-h predictions for the 36- and 12-km domains, and 24-h predictions (24–48 h) for the two 4-km domains. The model outputs are available every hour for display and analysis purposes and archived for validation and for other future research projects. All 92 possible days for the 2003 summer season (June through August) and a total of 79 out of 90 possible days for the winter season (December 2002 though February 2003) are available for verification. The missing-day predictions, all of which occurred in December of 2002, were due either to cluster downtime or to failure to access the Eta operational forecasts for model initialization.

3. Observational data and evaluation methods

The present validation of MM5 predictions is limited to the western 4-km domain over the Great Lakes regions as shown in Fig. 1. The outputs from both 12- and 4-km forecasts over this domain were compared to determine the effect of enhanced horizontal resolution in this region. The predictions with 36-km grid spacing were not examined because the two-way interactive nesting produced identical forecasts between the 12- and 36-km grids over this domain.

Data used for the validation in this domain came from several different sources, including hourly observations from approximately 193 U.S. Surface Airways stations, daily summaries from 669 Cooperative Observer (COOP) sites, and six twice-daily upper-air sounding sites. Figure 2 shows the observational sites used in the model verifications.

The MM5 results at grid points were interpolated to the irregularly spaced observational sites using a Cressman-type interpolation scheme (Cressman 1959). For any variable, ϕ, the modeled values at the four points surrounding the observational site are interpolated to the observational site following

$$\phi = \frac{\sum_{i=1,4} W_i \phi_i}{\sum_{i=1,4} \phi_i}, \tag{1}$$

where ϕ_i is the model value at the one of the four grid points, and W_i is the weight of the interpolation associated with ϕ_i, which is determined by

$$W_i = \frac{(\Delta x)^2 - r_i^2}{(\Delta x)^2 + r_i^2}, \tag{2}$$

where Δx *i*s the model horizontal grid spacing, and r_i is the horizontal distance between the *i*th surrounding point and the observational site.

To quantify forecast errors for near-surface temperature, humidity, wind speed, and wind direction, several standard statistical measures are computed. They include

bias:

$$\overline{\phi'} = \frac{1}{N}\sum_{i=1}^{N} \phi_i', \tag{3}$$

root-mean-square error:

$$\text{rmse} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(\phi_i')^2\right]^{1/2}, \tag{4}$$

and standard deviation of error:

$$\text{sde} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(\phi_i' - \overline{\phi'})^2\right]^{1/2}, \tag{5}$$

where $\phi' = \phi_{\text{mdl}} - \phi_{\text{obs}}$ is the departure of the modeled variables from the observed values. The errors include

contributions from systematic and nonsystematic sources. Systematic errors, represented by the bias, are usually caused by 1) consistent misrepresentation of local properties such as topography and land use, 2) physical mechanisms, such as cumulus convection and radiation, or 3) numerical factors. Nonsystematic errors, indicated by the error standard deviation, represent the random error components caused by uncertainties in model initial and boundary conditions or uncertainties in the observations.

A different set of statistical scores is computed to evaluate the skill of precipitation forecasting, including the bias score (BS), the threat score (TS), and the Heidke skill score (HSS). These statistical scores are defined based on a contingency table in which individual elements represent the number of events where the forecasted and observed precipitation amounts fall into certain threshold classes for a given forecast period.

The bias score is defined by

$$\mathrm{BS} = \frac{F}{O} = \frac{A + B}{A + C}, \tag{6}$$

where *A, B, C,* and *D* represent, for a given threshold class, correct positive, false positive, false negative, and correct negative, respectively. Here, $F = A + B$ and $O = A + C$ are the number of data points in which the forecasted or the observed precipitation amounts are within a given class, respectively. The bias score indicates how well the model predicts the frequency of occurrences of a given precipitation class. A perfect prediction should produce only *A* and *D*, which yields BS = 1. When averaged over many stations and some periods of time, BS < 1 or BS > 1 indicates that the model systematically underpredicted or overpredicted the occurrences of precipitation for a given threshold value.

The threat score is defined as

$$\mathrm{TS} = \frac{A}{A + B + C}, \tag{7}$$

which represents the ratio of the number of hits to the total number of occurrences in which the event was either predicted, or observed, or both. The TS score is very sensitive to the number of hits but is not influenced by correct negatives, or the number of cases in which the model correctly forecasted the nonoccurrence of a given precipitation class. It has been shown (Marzban 1998) that the lack of influence of correct negatives tends to produce biased TS scores in rare-event situations such as heavy precipitation events.

The Heidke score, defined as (Wilks 1995)

$$\mathrm{HSS} = \frac{2(AD - BC)}{(A + C)(C + D) + (A + B)(B + D)}, \tag{8}$$

measures the precipitation forecast skill in comparison to a particular standard, such as forecasts based on random chance. If the forecast is perfect, $B = C = 0$, leading to HSS = 1. Forecasts equivalent to random chance forecasts ($AD = BC$) yield HSS = 0, while worse-than-random-chance forecasts ($AD < BC$) produce a negative HSS score.

4. Results

a. Near-surface properties

Figures 3a and 3b show time series of predicted and observed domain mean and standard deviation of near-surface variables, including daily maximum and minimum temperature, daily averaged temperature, specific humidity, and wind speed, for the summer and winter seasons, respectively. The observed daily maximum and minimum temperature values were obtained using daily summary data from the 669 COOP sites in the validating domain, and the daily averaged temperature, humidity, and wind speed were computed using hourly observations at the 193 U.S. Surface Airways stations within the domain.

In the summer season, the forecasted daily maximum temperature values are consistently lower than the observed, and minimum temperatures are always higher. Consequently, the amplitudes of the diurnal temperature cycles are smaller in the model forecasts, but the daily mean temperatures show good agreement with the observations because of the cancellation of the errors in the predicted maximum and minimum temperatures. The smaller forecasted diurnal cycles are consistent with a wet bias of 1–2 g kg^{-1} near the model surface. Little bias is found in the forecasts of near-surface wind speeds.

Similar to the summer season, the winter forecasts of daily maximum temperature are systematically lower than the observed, but the minimum temperatures do not show a consistent warm bias as they do in the summer. The existence of a cold bias both day and night leads to a larger cold bias in the predicted daily mean temperature in winter than in the summer. Unlike the summer, the cold bias in the winter season can no longer be simply attributed to a more humid near-surface model atmosphere since the predicted humidity values appear to agree well with the observations. No systematic over- or underprediction is found in wind speed forecasting.

Several features are similar for the two seasons. First, there appears to be no clear correlation between the magnitudes of forecast errors for each variable and the magnitudes of the variable itself. For example, larger temperature forecast errors occurred during periods when observed temperatures were higher than the seasonal average, as well as periods when they were lower. The same can be said for wind and humidity forecast errors. Therefore, no conclusions can be drawn regarding whether the forecasted near-surface temperature, humidity, and winds are more accurate for a particular type of condition. Second, the standard deviations in the model forecasts are comparable to the standard de-

viations computed using the observations, suggesting that the forecasts captured the observed spatial variations of these near-surface variables. Finally, the forecasts adequately produced the observed day-to-day variations associated with changes in synoptic conditions, which is due primarily to the twice-daily initialization from observations that limit the error growth in the forecasting.

To quantify errors of near-surface variables, comparison statistics are computed using Eqs. (3)–(5) based on hourly observations at 193 U.S. Surface Airways stations. More than 42 600 data points for the summer and approximately 36 900 data points for the winter were used to compute the comparison statistics for temperature and wind; fewer data points were available for humidity. The results are summarized in Tables 1 and 2 for the 4- and 12-km domains, respectively. In the summer for the 4-km forecasts, the predicted temperature has a cold bias of −2.21°C during the day and a warm bias of 1.07°C at night, yielding a small overall bias of −0.57°C. The overall bias for the winter season is much larger at −2.67°C, resulting from a combination of a cold bias both day (−3.33°C) and night (−2.01°C). The humidity bias is positive during both the daytime and the nighttime, with an overall bias of 1.24 g kg^{-1} for the summer and 0.22 g kg^{-1} for the winter. For temperature and humidity, the biases and the error standard deviations are comparable in magnitudes, indicating that the systematic and nonsystematic errors contribute more or less equally to the total temperature and humidity errors. In both seasons, the wind speed bias is less than 0.5 m s^{-1}, and the direction bias is around 10°, which is much smaller than their corresponding error standard deviation values (1.7–2 m s^{-1} and 35°–55°). This indicates that nonsystematic errors comprised a large portion of the total errors in the predicted winds. The statistical measures computed using the 12-km resolution forecasts are very similar to those of 4-km forecasts, suggesting that a decrease in horizontal grid spacing from 12 to 4 km does not necessarily lead to an improvement in the accuracy of forecasts of near-surface temperature, humidity, wind speed, and wind direction.

The colder and wetter MM5 forecasts are consistent with the findings from several previous studies (Manning and Davis 1997; de Arellano et al. 2001; Zhong and Fast 2003, among others). The cold and wet biases have been attributed in the past to an inadequate specification of soil moisture in the model, which is also likely to be the dominant factor here because the summer of 2003 is known to have been unusually dry in the upper Midwest and Northeast. The soil moisture values in the model, specified based on land-use categories, would not be representative for these general dry conditions. The same explanation may not apply to the winter, when the cold bias is larger and wet bias is much smaller compared to the summer.

b. Precipitation

The MM5 precipitation results were compared to the observed precipitation by interpolating modeled precipitation amounts to the 669 COOP locations in the validation domain using Eqs. (1) and (2). The precipitation validation statistics were computed using Eqs. (6)–(8) for nine successive precipitation thresholds (0.25, 2.5, 5, 10, 15, 25, 35, 50, and 70 mm day^{-1}). The statistical scores were computed separately for the summer and winter seasons and for both 12- and 4-km forecasts. The results are shown in Fig. 4.

In the summer season, the bias scores for the 12-km forecasts increase slowly from just above 2 to about 4 for the light to moderate precipitation categories and rapidly to over 10 for the heavy precipitation category. This suggests that the 12-km grid substantially overpredicts the occurrences of precipitation for all categories, especially for heavy precipitation. The 4-km grid exhibits much better precipitation forecast skill with bias values ranging from 2 in low to moderate thresholds to approximately 3 at high thresholds. For both resolutions, the threat score and the Heidke score decrease rapidly with increased thresholds. These decreases, taken together with the increase in bias score, indicate a rapid decay of precipitation predictive skill from light to heavy precipitation. Similar to the summer, the winter season statistics show poor skill in heavy precipitation forecasting for both grids, especially for the 12-km grid. However, instead of a steady decay of skill from light to heavy precipitation as in the summer, the winter season shows a decrease in the bias and an increase in the HSS score when the threshold values increased from 0.25 to 5–10 mm day^{-1}, indicating improved skill from the light to moderate precipitation categories.

A better way to examine the precipitation forecast skill is to compare the frequency of the occurrences of the four individual elements that make up the statistical scores (Fig. 5). Recall, that the four elements, A, B, C, and D, represent, for a given threshold value, correct positive, false positive, false negative, and correct negative, respectively. For all thresholds, the frequency of false positive is much higher than that of false negative, suggesting that the model has a much higher chance to produce precipitation that is absent from the observation than to miss an observed precipitation event. It is interesting to note that the frequency of forecast hits or correct positives (A) is very similar between the 4- and 12-km forecasts, although the latter exhibits much higher bias scores. The higher 12-km bias scores, defined by BS = $(A + B)/(A + C)$, are attributed to the higher number of false positive (B) compared to false negative values (C) as shown in Fig. 5. The rapid decrease of A, B, and C values at higher precipitation thresholds with $A \rightarrow 0$ and $B >> C$ explains the large bias scores and near-zero threat scores in the categories of heavy precipitation.

The contingency table–based evaluation statistics, al-

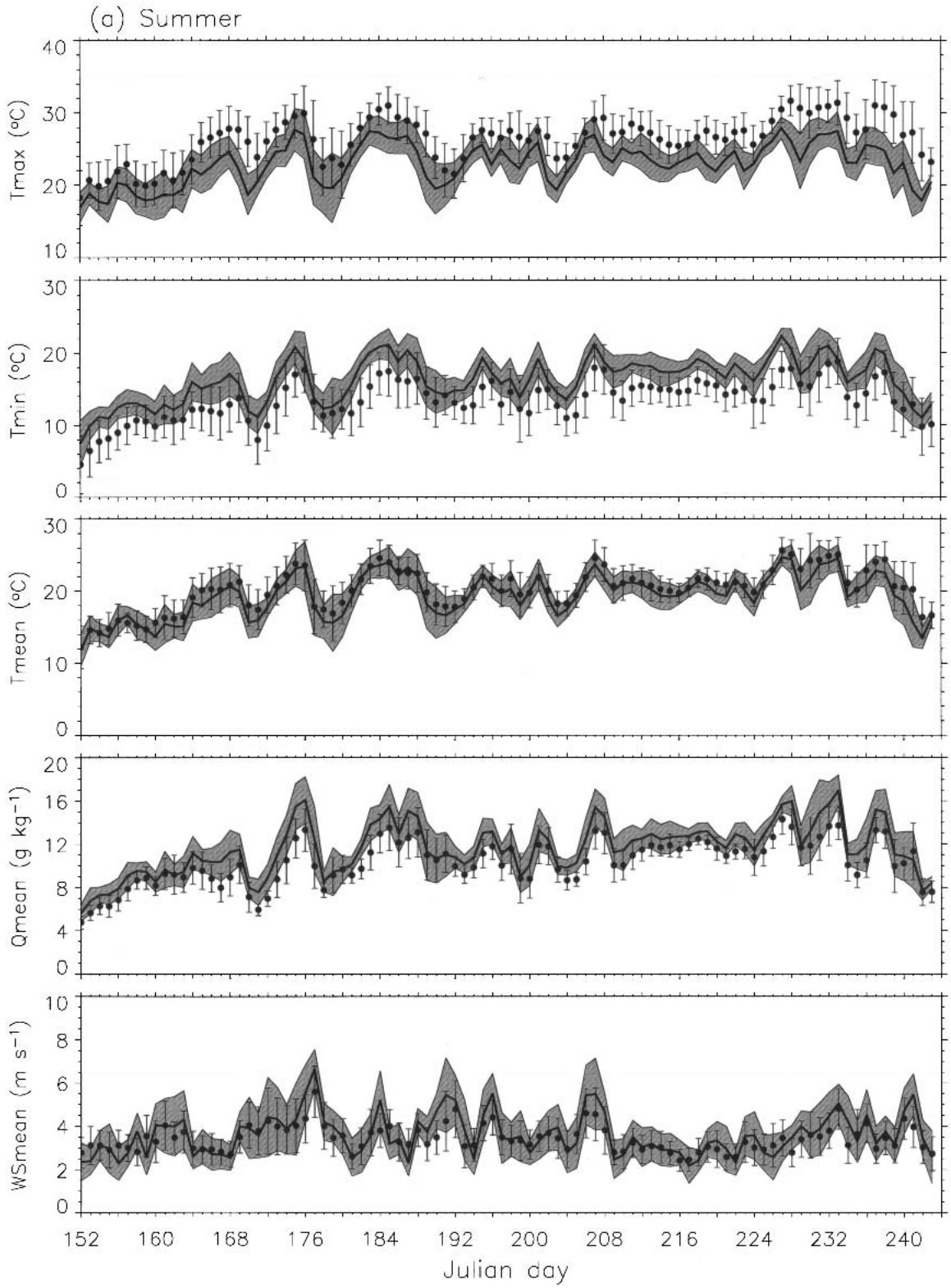

FIG. 3. Time series of observed and forecasted domain mean values and standard deviations of daily maximum temperature, daily minimum temperature, daily mean temperature, specific humidity, and wind speed for the (a) summer season and (b) winter season. The predicted domain mean values and standard deviations are represented by the line and shading, respectively. The observed domain mean values and the standard deviations are denoted by the circles and the bars, respectively.

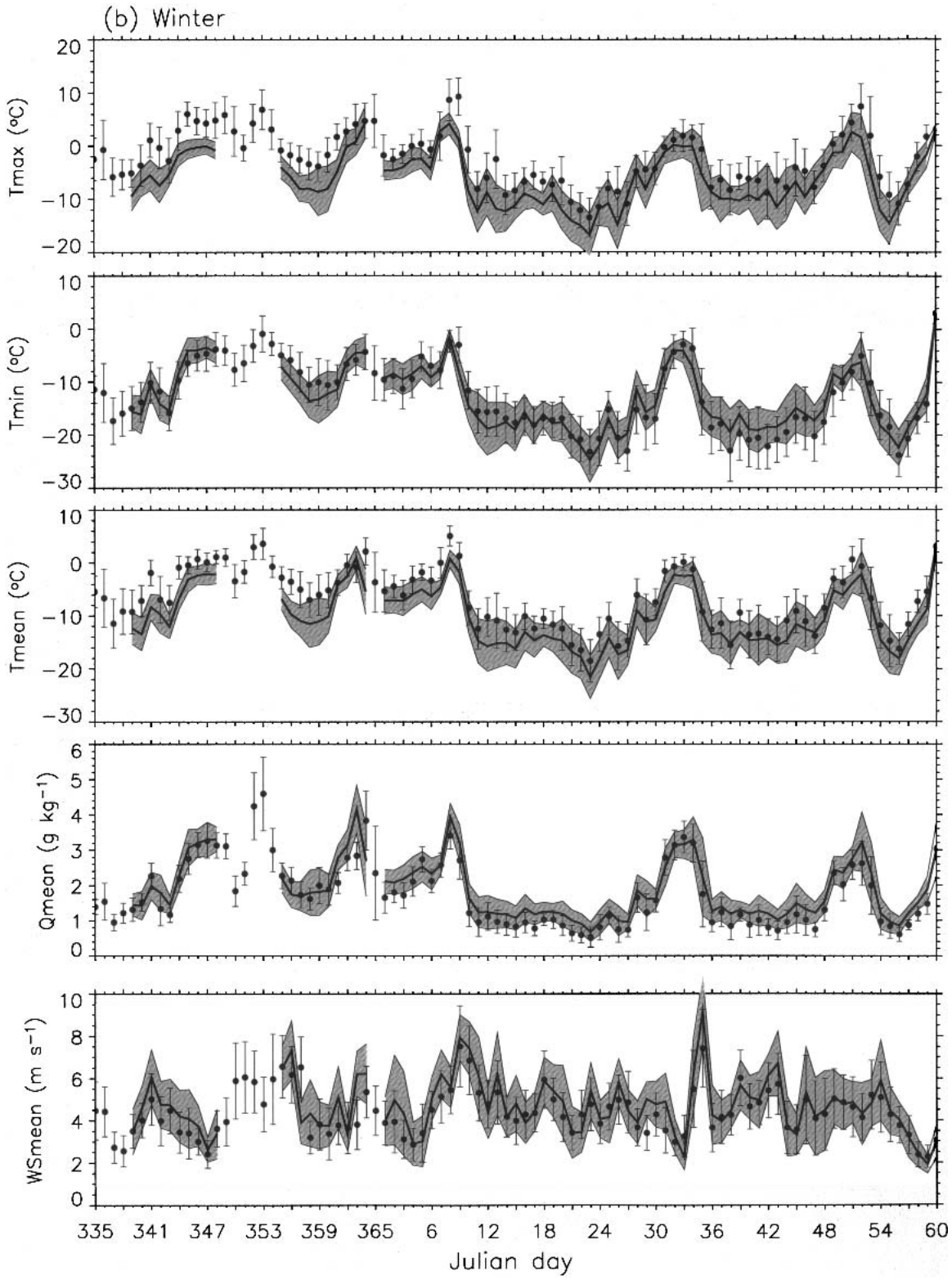

FIG. 3. (*Continued*)

TABLE 1. Bias, root mean square error, and standard deviation of error (SDE) for near-surface temperature (T), specific humidity (Q), wind speed (WS), and wind direction (WD) calculated for the 12-km domain for the summer and winter seasons and for daytime and nighttime.

		Summer			Winter		
		Bias	Rmse	SDE	Bias	Rmse	SDE
T	All	−0.63	3.07	2.27	−2.70	4.09	2.72
	Day	−2.29	3.53	2.40	−3.37	4.54	2.75
	Night	1.03	3.53	2.13	−2.03	4.54	2.70
Q	All	1.24	2.01	1.38	0.22	0.48	0.36
	Day	0.84	1.83	1.45	0.17	0.48	0.37
	Night	1.65	1.83	1.32	0.26	0.48	0.36
WS	All	0.36	1.75	1.57	0.50	1.93	1.72
	Day	0.08	1.79	1.67	0.29	1.97	1.78
	Night	0.65	1.79	1.47	0.70	1.97	1.67
WD	All	10.9	53.5	51.0	9.57	35.5	32.4
	Day	9.11	53.7	51.5	9.37	35.3	32.1
	Night	12.6	53.7	50.5	9.77	35.3	32.7

TABLE 2. Same as in Table 1, but for the 4-km domain.

		Summer			Winter		
		Bias	Rmse	SDE	Bias	Rmse	SDE
T	All	−0.57	3.01	2.22	−2.67	4.06	2.72
	Day	−2.21	3.42	2.33	−3.33	4.50	2.74
	Night	1.07	3.42	2.11	−2.01	4.50	2.70
Q	All	1.23	2.02	1.40	0.22	0.48	0.36
	Day	0.84	1.85	1.46	0.17	0.47	0.37
	Night	1.63	1.85	1.33	0.26	0.47	0.35
WS	All	0.40	1.76	1.59	0.50	1.95	1.74
	Day	0.13	1.80	1.68	0.30	1.99	1.80
	Night	0.67	1.80	1.49	0.70	1.99	1.69
WD	All	11.1	53.8	51.3	9.39	35.6	32.7
	Day	9.26	54.2	51.9	9.25	35.5	32.4
	Night	12.9	54.2	50.8	9.53	35.5	33.0

though useful in measuring the model skill, provide little information about the actual errors in the precipitation prediction for a given time period. To understand the difference between the observed and pre-

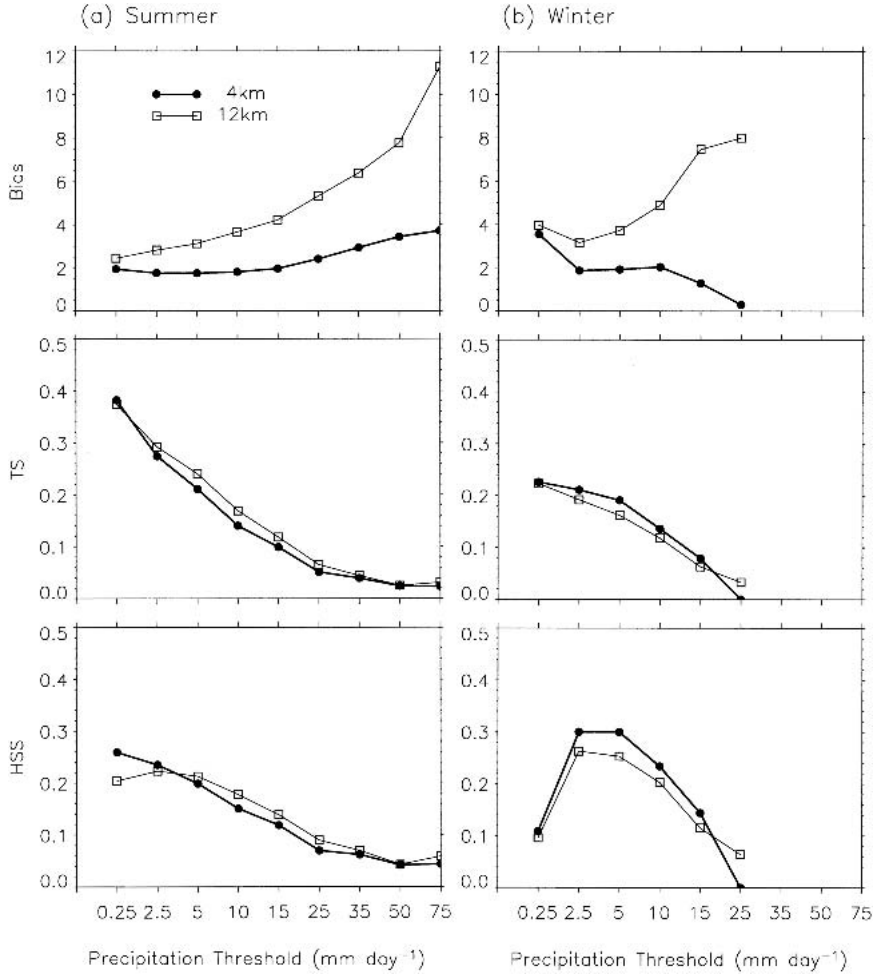

FIG. 4. The bias, threat, and Heidke precipitation forecasting scores for the (a) summer season and (b) winter season.

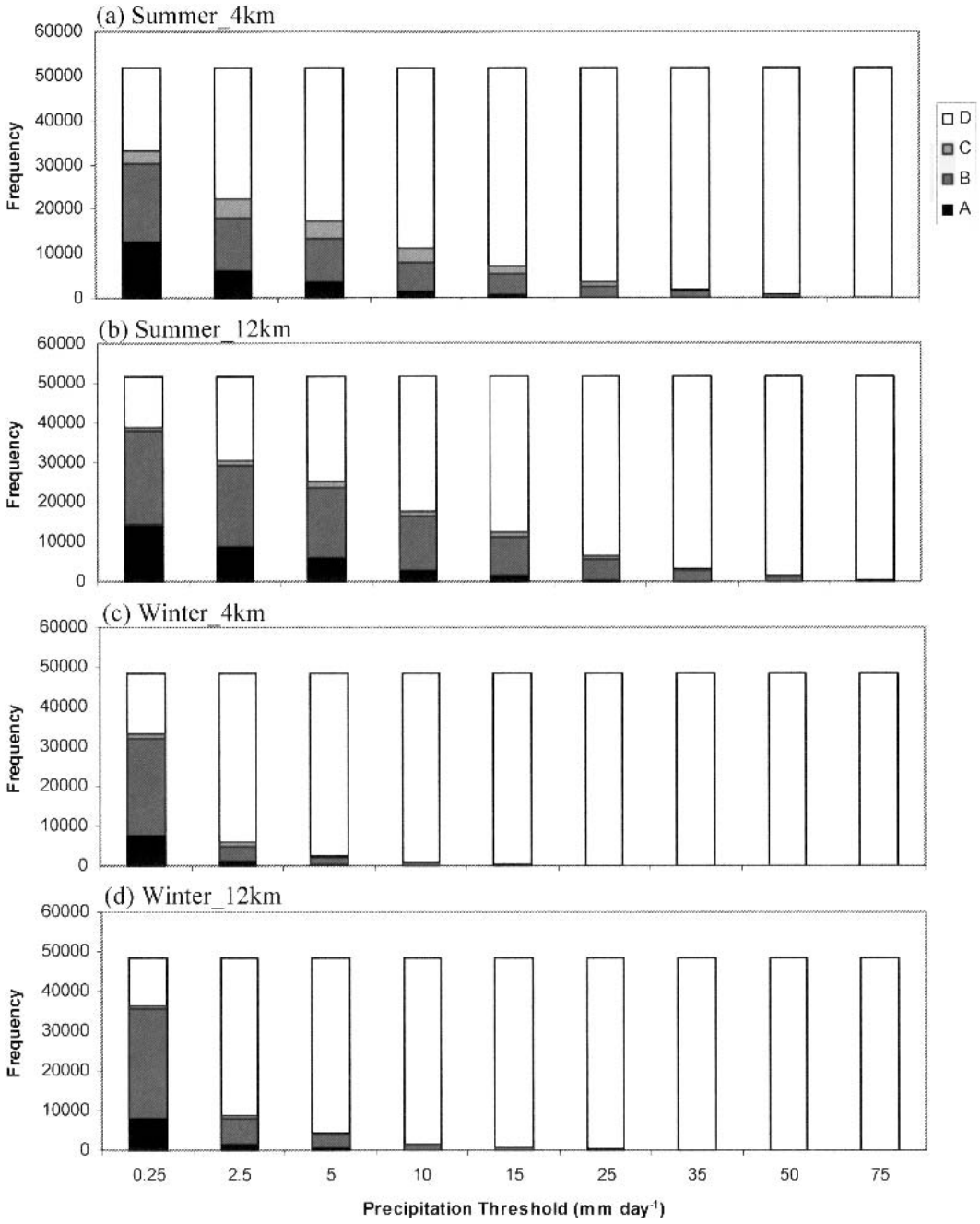

FIG. 5. The frequency of occurrences of the four individual elements used for computing the precipitation scores.

dicted precipitation amount, the predicted total precipitation amount for each of the summer and winter months is compared to the observations. The results are shown in Fig. 6. The model consistently overpredicted precipitation amounts for all months, and the overprediction is much worse in the 12-km results than in the 4-km results. The monthly total precipitation amount predicted with the 4-km grid spacing is 1.3–2 times higher than the observed amounts, compared to 2–4 times higher in the 12-km predictions.

To understand whether these errors were the result of a small amount of overprediction spread throughout the season, or whether they came from a large amount of overprediction during a few isolated events, time series of domain-accumulated daily total precipitation are compared with the observed values. The comparisons are shown in Figs. 7a and 7b for the summer and winter seasons, respectively. Both plots show that the forecasted precipitation amounts were always higher than the observed except for a few days in the summer.

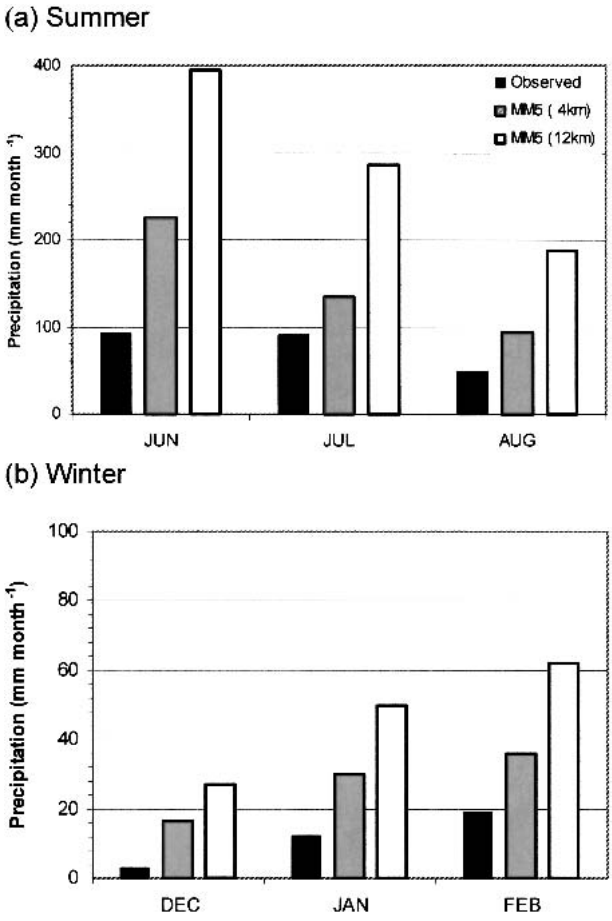

FIG. 6. Observed and forecasted monthly domain total precipitation for (a) summer and (b) winter months.

The 12- and 4-km forecasts tend to closely track each other with the 12-km forecasted precipitation amount being persistently higher. The large error in the monthly total precipitation in June, as seen in Fig. 6, resulted primarily from the large overforecasting of the most heavy precipitation event in the summer that occurred on 23–26 June (Julian days 174–177). A similar explanation may be given to the large error in February when the 12-km predicted precipitation nearly tripled the observed amount during a heavy precipitation event that occurred on 3–4 February (Julian days 34–35). In both seasons, the forecasts appear to track the day-to-day changes of the observed precipitation reasonably well.

The overprediction in domain total precipitation at a given time can be a result of larger area coverage and/or substantially larger amount at some locations in the forecasts. To isolate these two factors, the time series of the predicted domain-accumulated daily total precipitation in Figs. 7a and 7b, which was computed using predicted precipitation at all 699 COOP sites regardless of whether precipitation was observed at the sites, is replaced by a new time series obtained using only those sites at which precipitation occurred (Figs. 8a and 8b). The predicted precipitation amount is now in very good agreement with the observations in both summer and winter seasons, indicating that the overprediction seen in Figs. 7a and 7b is due to more locations with predicted precipitation rather than excessive predicted precipitation at individual locations.

The more widespread area coverage in the precipitation prediction is also evident in the comparison of the predicted and observed area coverage of monthly total precipitation shown in Fig. 9 for June 2002 and February 2003. In February, less precipitation occurred in north-central Minnesota and southern Wisconsin and more occurred in Michigan, especially in the southeastern part of the state and in the Upper Peninsula along the southern boundary of Lake Superior. The forecasts captured this overall pattern reasonably well, but the predicted area coverage was larger than the area coverage in the observation. In June, more precipitation occurred in the western part of the domain than the eastern part, with the heaviest amount in east-central Minnesota and northern Iowa and the least amount in southern Wisconsin and most of Michigan. Although the forecast captured this general spatial distribution pattern, the heavy precipitation with amounts exceeding 140 mm month^{-1} appears to spread over a larger area in the forecasts than in the observations.

To examine the sensitivity of predicted summertime precipitation to convective parameterizations in the model, a 10-day (26 July–5 August 2003) forecast was repeated with the Kain–Fritsch cumulus convection scheme employed in the original run replaced by the Grell scheme (Grell 1993). In the Grell cumulus parameterization, clouds are maintained by the updraft and downdraft originated at levels of maximum and minimum ambient moist static energy with no mixing between the cloud and environment except at the top and bottom of the circulations. Convection is triggered when a parcel lifted from the updraft originating level attains moist convection as determined by moist static energy. The Kain–Fritsch scheme uses a classical approach to remove convective available potential energy (CAPE) by vertical reorganization of mass. This method includes detrainment of mass and moisture from deep convective clouds and represents the exchange of mass between cloud and environment. Convection is triggered based on parcel buoyancy at the lifting condensation level. Because of differences in closure assumptions and trigger functions, large differences can often be found in the amount and spatial distribution of precipitation simulated by different convection schemes.

Figure 10 shows a comparison between domain-accumulated daily total precipitation using the two schemes with the observed precipitation for the 10-day period. Both schemes overpredicted precipitation on each of the 10 days with the Grell scheme being consistently higher. Since the forecasts were driven by the

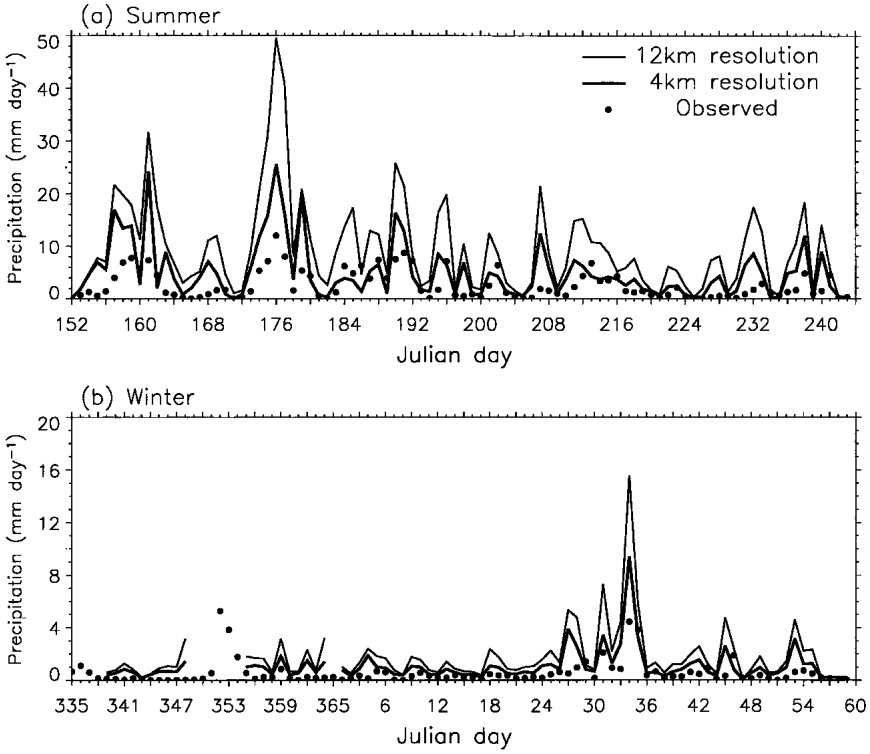

FIG. 7. Observed and forecasted domain total precipitation as a function of days computed using data from all 699 COOP sites for the (a) summer season and (b) winter season.

same large-scale conditions, the difference in the precipitation between the two schemes is likely to be caused by the difference in the convection-triggering mechanisms. Similar spatial distributions (not shown) were produced by the two schemes over most of the domain except near the western boundary where the Grell scheme produced significantly more precipitation. This sensitivity test, although limited to a short

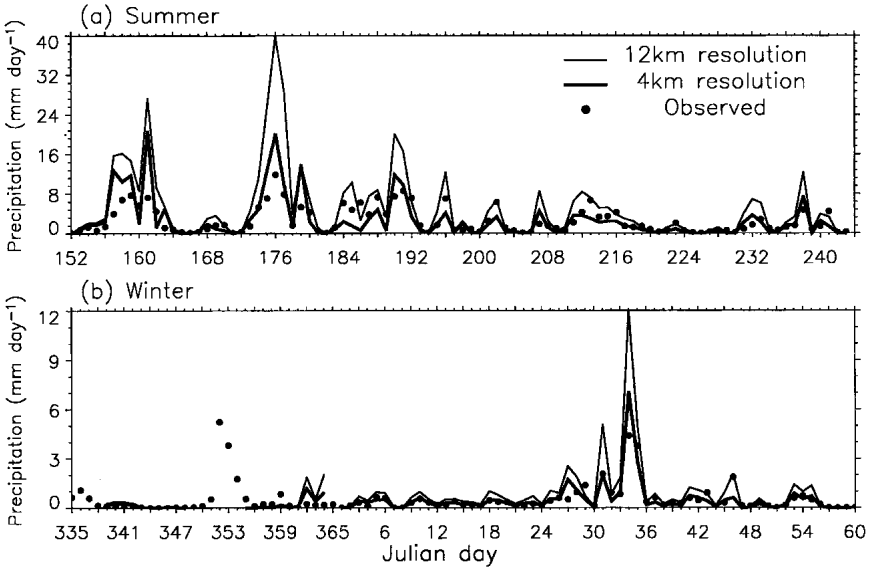

FIG. 8. Observed and forecasted domain total precipitation as a function of days computed using data from the COOP sites at which there was observed precipitation for the (a) summer season and (b) winter season.

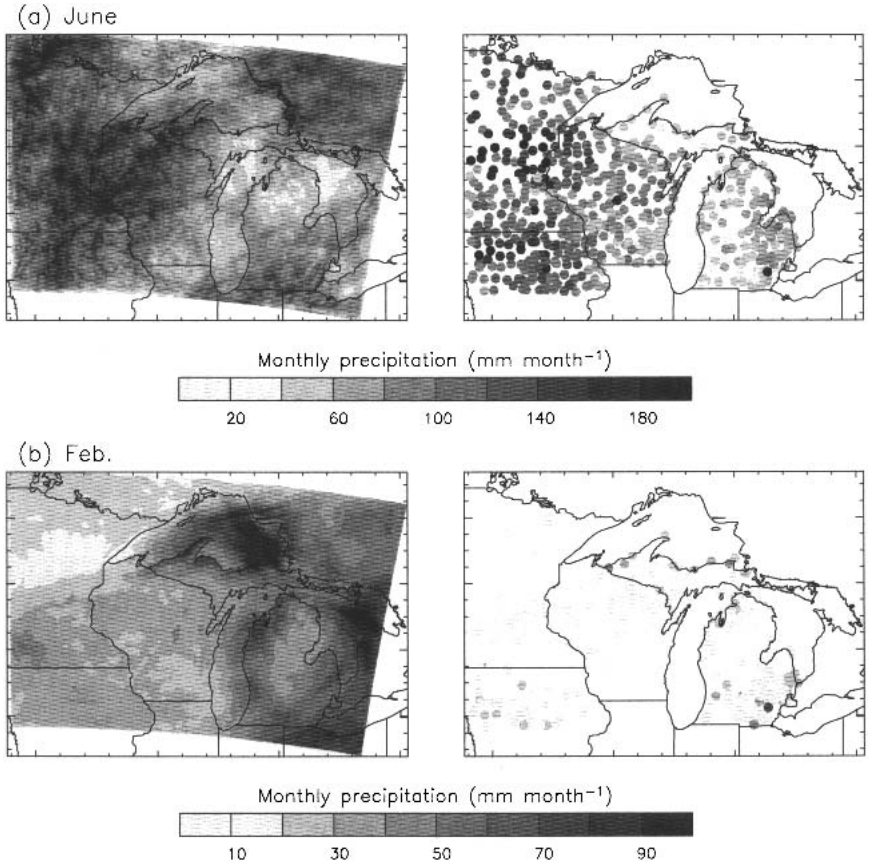

FIG. 9. Observed and forecasted spatial distribution of precipitation for (a) Jun and (b) Feb.

time period, suggests that the summer season precipitation forecasts for the current domain of the Great Lakes region are unlikely to be improved by using the Grell scheme.

c. Vertical soundings and boundary layer structure

To understand how well the model predictions were able to capture the observed vertical structure of the atmosphere, the predicted vertical profiles of potential temperature, specific humidity, and wind averaged for the summer and the winter seasons, respectively, are compared to the corresponding mean vertical profiles available twice per day at 0000 and 1200 UTC at the six upper-air sounding sites in the validating domain. The comparisons yielded similar results among the six sites, and the results are, therefore, shown in Fig. 11 for only three of the six sites.

For the mean temperature, the predicted values are lower than the observed, especially in the lowest 1500 m in the summer and 500 m in the winter, which corresponded to the average depth of the boundary layer in the two seasons. The values of the cold bias in the lower atmosphere are 2°–3°C in the summer and 1°–2°C in the winter. The cold bias is more pronounced in the afternoon (0000 UTC) than in the early morning (1200 UTC). The predicted surface-based inversion at 1200 UTC appears to be much weaker than observed in the summer season, becoming somewhat stronger than observed during winter. The predicted specific humidity profiles are wetter by 0.5–1.5 g kg^{-1} in the boundary layer and drier above the boundary layer. The errors in the predicted wind profile appear to be larger in the winter than in the summer, and the departure of the predicted and observed wintertime winds increases with height, with winds aloft being predominantly westerly in the forecast compared to west-northwesterly in the observation. The errors in the two wind components, however, tend to cancel each other to produce total wind speeds that are in good agreement with the observed wind speeds. Some of these findings, such as the cold bias in the boundary layer, are very similar to what was found by Zhong and Fast (2003) in their comparisons of MM5 simulations with soundings taken in the Salt Lake Valley during a field campaign, although there are large differences in topography and land use between Salt Lake Valley and the Great Lakes region.

Among the various atmospheric properties produced in mesoscale forecasting, one that is important for

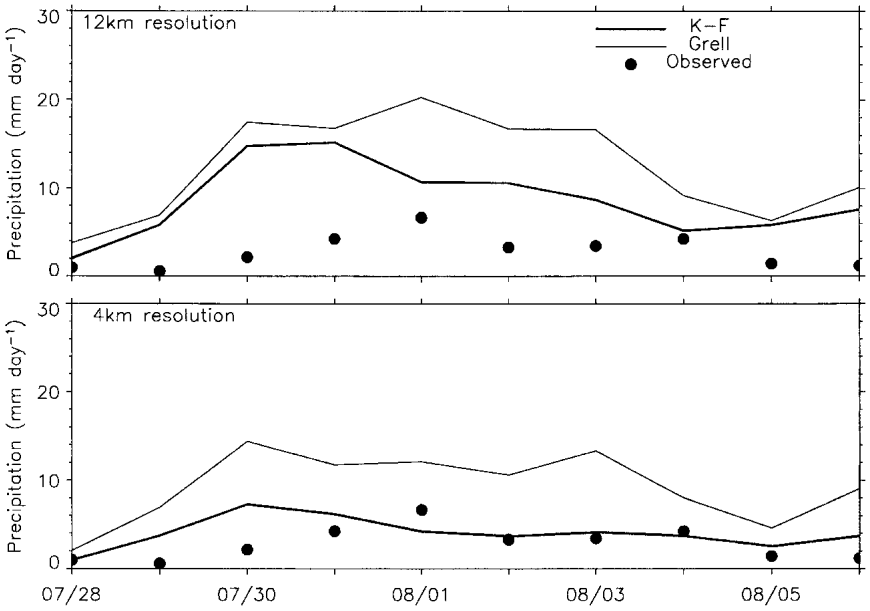

FIG. 10. Comparison of observed (dots) and predicted precipitation by the Kain–Fritsch convective parameterization scheme (thick line) and the Grell scheme (thin line) for a 10-day period from 28 Jul to 5 Aug for the (top) 12- and (bottom) 4-km resolution domains.

many applications is the mixed-layer height. For example, as an input variable to air quality models, the depth of the mixed layer and the rate at which it grows have a direct impact on the mixing, transport, and dispersion of pollutants. To evaluate how well the MM5 forecasts capture the afternoon mixed-layer depths, the modeled mixed-layer depths were compared to the observations at all six upper-air sounding sites in the validation domain. The mixed-layer depths are determined as the base of the elevated inversion using predicted and observed potential temperature profiles at 0000 UTC for the summer season. The results are shown in Fig. 12 for three of the six sites. The predicted mixed-layer depths exhibit no noticeable differences among the locations, consistent with the observations. However, at all locations, the predicted afternoon mixed-layer depths are substantially lower than those observed. This underprediction is consistent with results from earlier studies (among others, Berg and Zhong 2004, manuscript submitted to *J. Appl. Meteor.*) showing that the MM5 predicted mixed-layer depths are very sensitive to boundary layer turbulence parameterization schemes. These previous studies focusing on a few cases concluded that boundary layer parameterization schemes in which turbulent mixing is determined based upon the predicted turbulent kinetic energy (TKE), such as the Eta boundary layer scheme employed by the current real-time predictive system, tend to underpredict mixed-layer depths. The current results support this previous finding by extending the comparison from a few cases to an entire season.

Another important boundary layer property affecting atmospheric near-surface dispersion is the surface-based radiation inversion at night. Not only are nocturnal inversions a key meteorological factor in air pollution, but they also have a great impact on agriculture and aviation because of their role in the buildup of cold air pools in lower-lying terrain and the formation of fog and frost. The MM5-predicted inversion strengths in early morning were evaluated using the 1200 UTC soundings, and the results are shown in Fig. 13. The inversion strengths were determined by the potential temperature gradients between 10 and 200 m above ground. The sounding profiles were interpolated linearly to these levels before the gradients were computed. The data points from the three sites are mixed together, indicating small spatial variation of morning inversion strengths in both the forecasts and the observations. Except for a few data points, the predicted summer inversion strengths are much weaker than those observed. It is interesting to note that in the summer, while the observed values are all positive, indicating an increase of potential temperature with height in the lowest 200 m, roughly 20% of the predicted values are below zero. These negative values represent a decrease of potential temperature with height, suggesting that a mixed layer had already developed and grown to 200 m by 1200 UTC. An examination of individual potential temperature profiles corresponding to those negative data points confirmed that a mixed layer had indeed developed in the MM5 forecasts by 1200 UTC (0700 EST), just 1–2 h after sunrise, that occurred at 0458 EST on 1 June and 0556 EST on 31 August. Figure 14 shows several of these profiles at the three

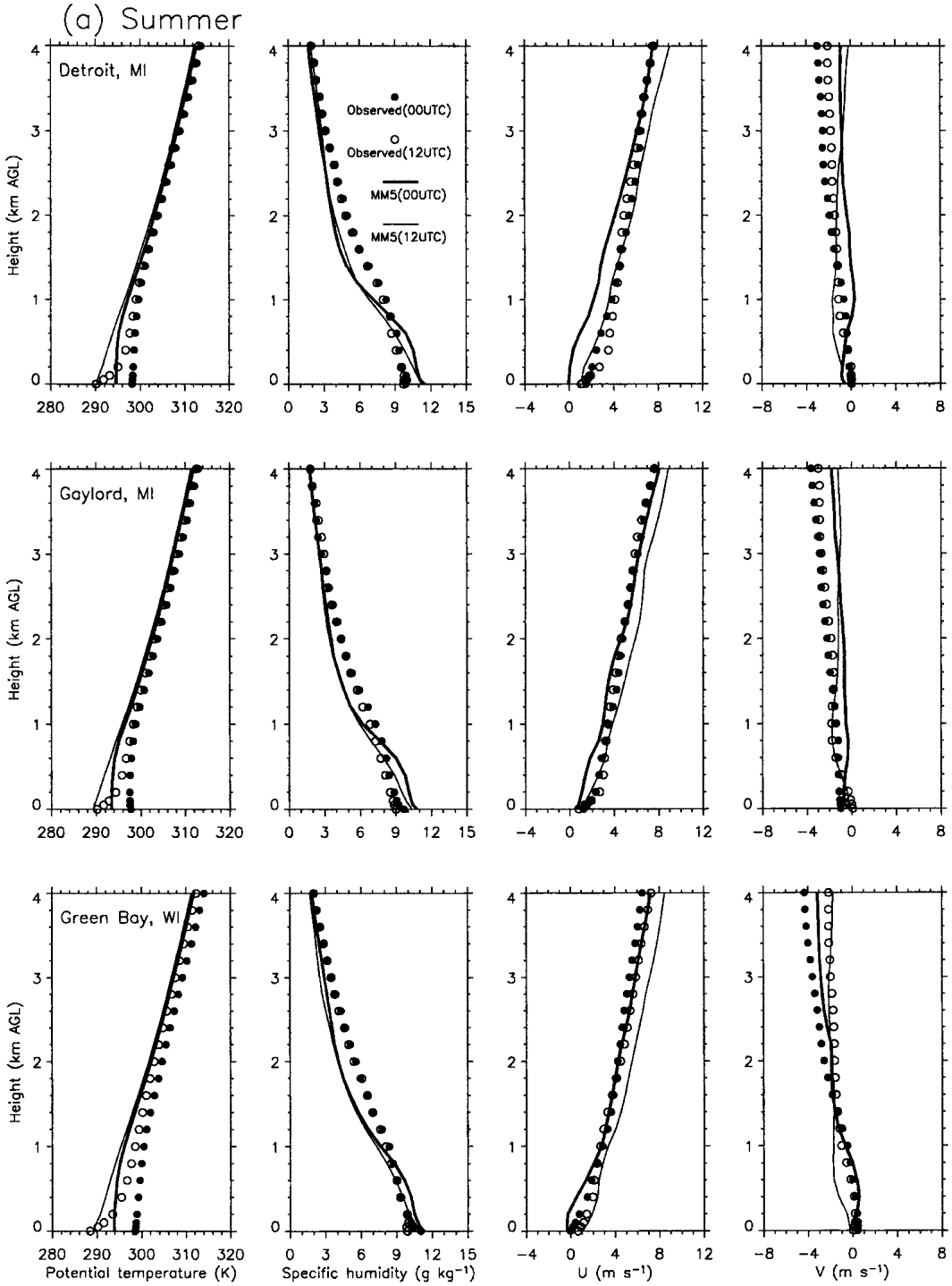

FIG. 11. Observed and forecasted mean vertical profiles of potential temperature, specific humidity, and u and v components of horizontal winds at 0000 and 1200 UTC at Detroit and Gaylord, MI, and Green Bay, WI, for (a) summer and (b) winter.

sites where a mixed layer had developed in the lowest 100–200 m in the forecast while the nocturnal inversion still existed in the observation. The earlier development of a mixed layer in the forecasts is consistent with the fact that the predicted nocturnal inversion is significantly weaker than the observed, and consequently, less solar heating is required to break up the inversion before a mixed layer starts to grow. The head start of

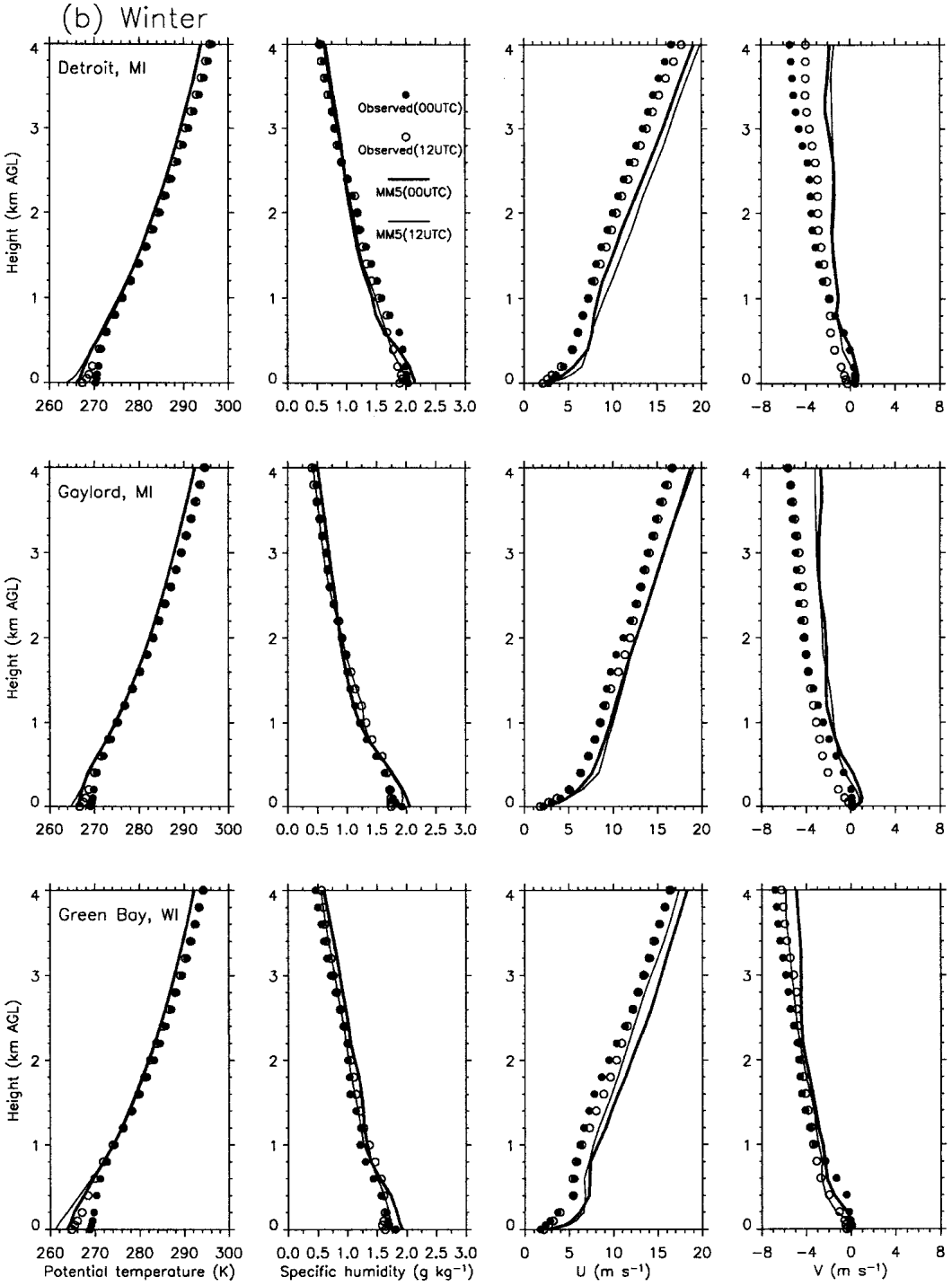

FIG. 11. (*Continued*)

the mixed-layer growth in early morning, however, failed to produce either a deeper afternoon mixed layer or a warmer convective boundary layer, suggesting that either the surface sensible heat flux in the model is lower than that observed or the flux divergence in the boundary layer may be too high. Unfortunately, no energy flux data are available to verify these hypotheses.

d. Lake breeze

The large lakes, complex shorelines, and resulting lake breezes are unique features of the Great Lakes

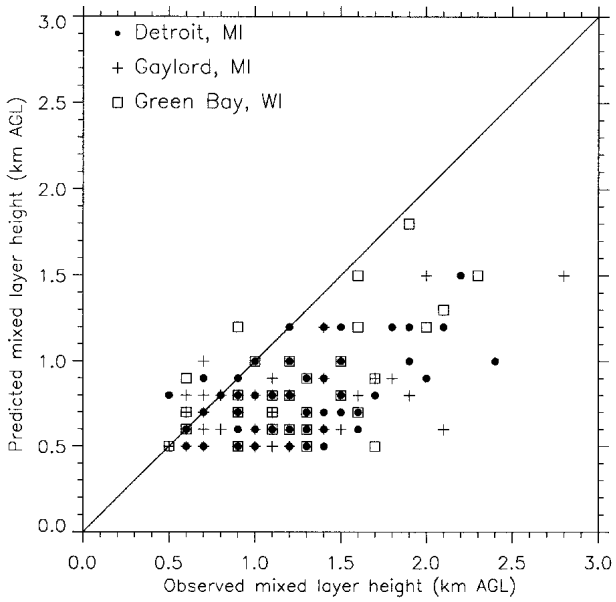

FIG. 12. Observed and forecasted mixed-layer heights at 0000 UTC over Detroit, Gaylord, and Green Bay for the summer season.

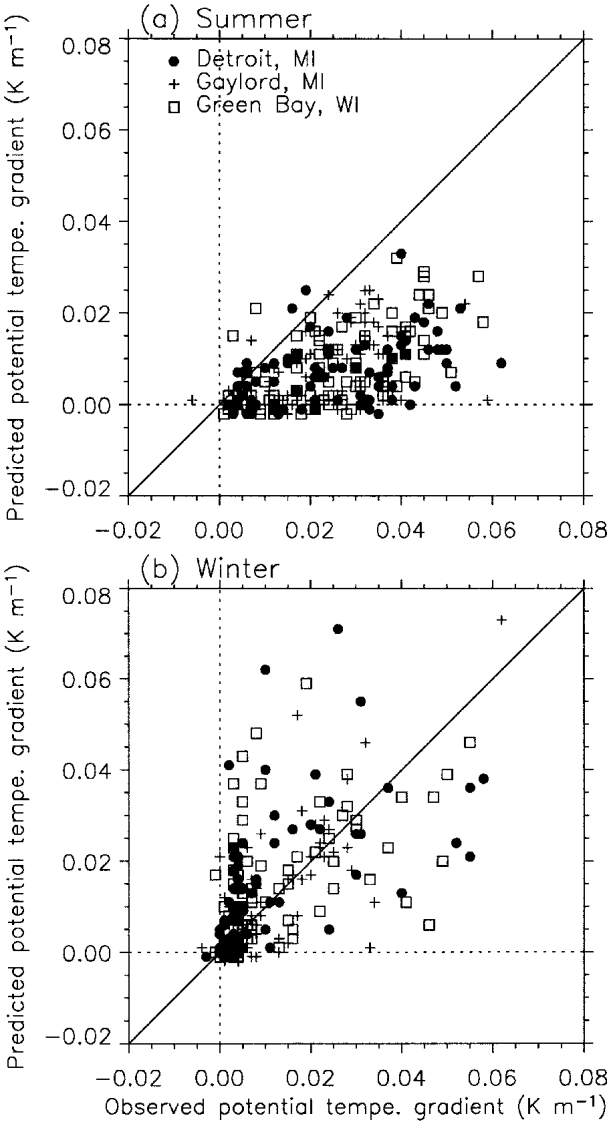

FIG. 13. Observed and forecasted 1200 UTC potential temperature gradients in the lowest 200 m over Detroit, Gaylord, and Green Bay for (a) summer and (b) winter.

region. To evaluate the ability of MM5 real-time forecasts in predicting the development and evolution of the lake–land breezes, we chose to focus on six surface sites surrounding Lake Michigan on days with weak synoptic winds and clear skies when these thermally driven circulations are expected to be well-developed. The soundings from Detroit, Michigan, on the east side of Lake Michigan and Green Bay, Wisconsin, on the west side are employed to determine the weak synoptic wind days using an arbitrarily chosen criterion that observed winds at 850 mb were less than 5 kt. The days that satisfy this criterion are the same regardless of which of the two sounding sites were used, and a total of 31 days were selected from the three summer months as having weak synoptic winds. For each of the 31 days, the station cloud cover information was used to further select days with clear skies using the criterion that low- and midlevel clouds should not be present at the location. Hourly near-surface winds from days that satisfy both criteria, which range from 10 to 23 days among the six locations, were used to construct hourly mean wind vectors for a diurnal cycle obtained by vector averaging the wind at each hour over the selected days. The resulting hodographs are shown in Fig. 15 for the six sites that are almost equally distributed on the eastern and western shoreline of Lake Michigan. The hodographs in Fig. 15 are deviations of the mean wind vectors for each of the 24 h of the day from the 24-h average wind to better illustrate the diurnal nature of the winds.

The turning of wind vectors, clockwise on the eastern side and counterclockwise on the western side of Lake Michigan, is a clear indication of the development of a lake breeze in the afternoon and land breeze during night and early morning hours. The lake breezes appear to be somewhat stronger than the land breezes, especially on the eastern side of the lake, as indicated by larger vectors from the origin to the individual points on the hodograph in the afternoon hours as compared to the nighttime and morning vectors. The hodographs appear to take a more circular shape on the east side of the lake than the west, indicating that the lake and land breezes are better developed on the east side. At all sites, the model captured the general pattern of wind rotation quite well, but the predicted timing of the land to lake breeze transition and vice versa do not exactly match the observations. At some locations the modeled lake breeze appears to be slightly weaker during its peak hours in late afternoon compared to the

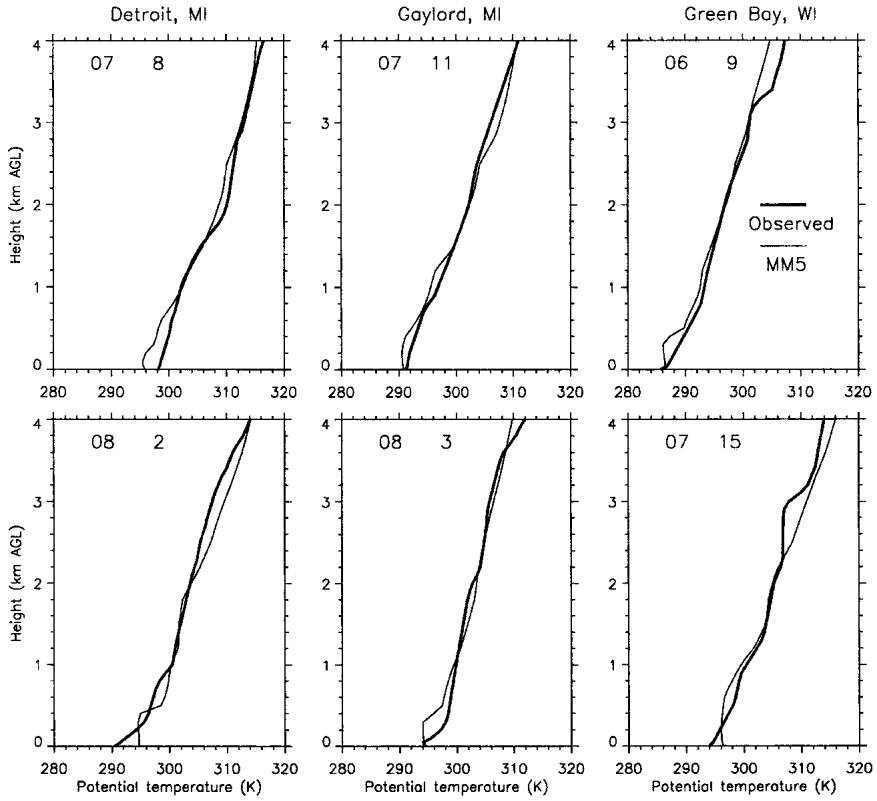

FIG. 14. Examples of 1200 UTC potential temperature profiles showing that a mixed layer has begun to develop in the forecasts while an inversion layer is still present in the observation.

observed speed. Overall, the model is able to predict the general development of lake–land breezes in the area surrounding the lake during the summer, although errors exist in the predicted strengths of the breezes as well as the timing of the transitions. Since the number of days satisfying our criteria is rather small in one summer, more data are needed to understand the source for the errors in lake–land breeze predictions.

5. Summary and conclusions

This study presents a detailed evaluation of real-time, operational high-resolution MM5 predictions for the Great Lakes region. The validation used surface and upper-air observations from various sources in the region and was performed for the winter season of 2002/03 and summer season of 2003. Although the reinitialization on each day allowed the model to capture day-to-day variations associated with changes in synoptic conditions, the forecasts fail to produce the details in spatial and time variation of the observed variables and contain errors that can be relatively large at times in the lower atmosphere.

In the summer season, the model exhibited a cold bias of 1°–3°C in the predicted daily maximum temperature and a warm bias of approximately 1°C in the minimum temperature. Consequently, the amplitudes of the diurnal oscillation of the predicted temperature are considerably smaller, but the predicted daily mean temperatures are in good agreement with the observations due to the cancellation of the daytime and nighttime errors. In winter, a cold bias of 2°–3°C is found during the daytime and 1°–2°C during the nighttime, leading to a larger cold bias of approximately 2°C in daily mean temperature but a good prediction of temperature diurnal cycle. The cold bias is not confined to near surface, but occurs in the entire boundary layer. The cold bias in the summer season may be explained by a wetter predicted boundary layer that, in turn, can be attributed to the climatologically based soil moisture value used in model runs, which was an inadequate representation for a particularly dry summer. The same explanation may not simply apply to the winter season forecast when the cold bias is larger, while the wet bias is much less compared to the summer.

For all precipitation categories and in both summer and winter, the model produced substantially more precipitation than was observed, especially in the heavy precipitation categories for which the model showed poor predictive skill. The model has a substantially higher chance of producing precipitation when it is absent in the observation at a particular location than of

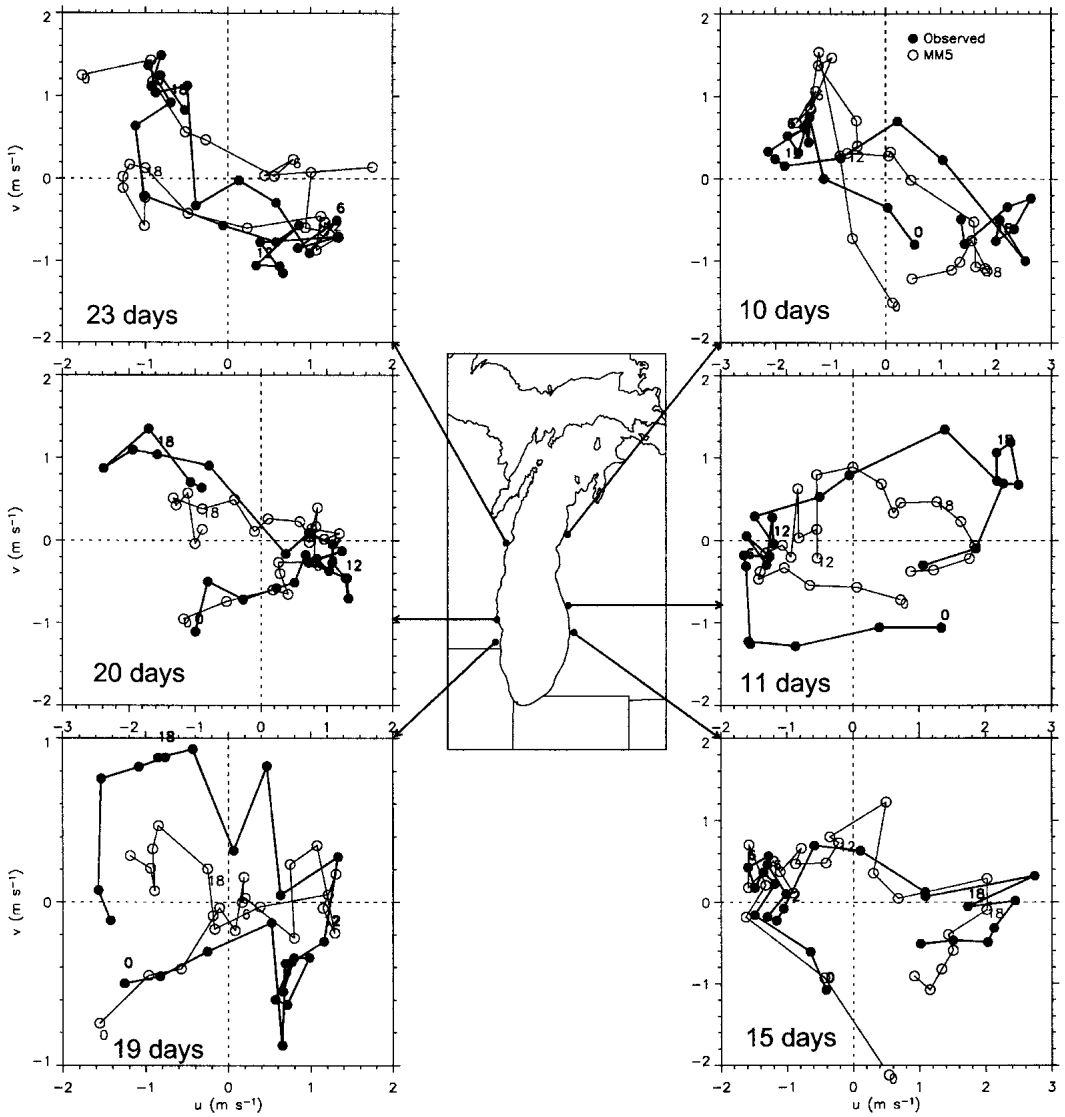

FIG. 15. Hodographs of surface winds at six locations on the shore of Lake Michigan. The hodographs were constructed by subtracting the 24-h mean wind vector from the wind vector at each of the 24 h of the mean summer day with light winds and clear skies. The filled circles represent the observations, and the open circles represent the forecasts. The number of days used in the composite at each site is indicated in the figure.

missing an observed event. It appears that the overprediction results primarily from more widespread area coverage rather than larger amounts in the prediction. The colder and wetter forecasts of the lower atmosphere compared to observations imply that the conduciveness of the atmosphere to extreme fire behavior is likely to be underestimated when the MM5 products are used to produce fire weather indices.

In both seasons, there are few systematic errors in the predicted near-surface and boundary layer winds, but the predicted winds aloft are generally from the west, while the observed winds are from the west-northwest. The model appears to be able to capture the development and evolution of lake and land breezes during the summer season, although the strengths of the lake–land breezes and the timing of their transition may not exactly match the observations.

The predicted afternoon mixed-layer depths are considerably lower than those observed. The predicted inversion strengths in early morning are significantly weaker than the observed inversion in the summer but somewhat stronger in the winter. The weaker surface-based inversion in the summer leads to a more rapid breakup of the inversion followed by an earlier development of a mixed layer in the morning forecasts. The forecasted head start of the mixed-layer growth in the morning, however, fails to produce a deeper mixed layer in the afternoon, suggesting that either the surface

sensible heat flux in the model may be too small, or there is too much flux divergence across the boundary layer. Given these relatively large errors in the modeled mixed-layer development and nocturnal inversion strengths, precautions need to be taken when these properties are used in applications such as air pollution forecasting.

Finally, there is little difference between the 12- and 4-km results in almost all the properties except for precipitation, for which the decrease of grid spacing from 12 to 4 km significantly reduced the bias of overprediction in all categories.

Acknowledgments. Hee-Jin In was supported partially by the USDA Forest Service under Research Joint Venture Agreement 03-JV-11231300-087 and by EPA Grant R-82906801.

REFERENCES

Aves, S. L., W. A. Gallus Jr., E. Kalnay, and M. Miller, 2002: The use of a phase shifted verification score to evaluate warm season QPF. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., JP2.7.

Bukovsky, M. S., P. Janish, J. S. Kain, and M. E. Baldwin, 2002: Evaluation of Eta model forecasts of mesoscale convective systems. Preprints, *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., JP3.2.

Chen, S., J. E. Nachamkin, J. M. Schmidt, and C.-S. Liou, 2002: Quantitative precipitation forecast for the coupled ocean/atmosphere mesoscale prediction system (COAMPS). Preprints, *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 7B.1

Chien, F.-C., Y. H. Kuo, and M.-J. Yang, 2002: Precipitation forecast of MM5 in the Taiwan area during the 1998 Mei-yu season. *Wea. Forecasting,* **17,** 739–754.

Colle, B., K. J. Westrick, and C. Mass, 1999: Evaluation of MM5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. *Wea. Forecasting,* **14,** 137–153.

——, C. F. Mass, and K. J. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest during the 1997–99 seasons. *Wea. Forecasting,* **15,** 730–744.

——, J. B. Olson, and J. S. Tongue, 2003a: Multiseason verification of the MM5. Part I: Comparison with the Eta model over the central and eastern United States and impact of MM5 resolution. *Wea. Forecasting,* **18,** 431–457.

——, ——, and ——, 2003b: Multiseason verification of the MM5. Part II: Evaluation of high-resolution precipitation forecasts over the northeastern United States. *Wea. Forecasting,* **18,** 458–480.

Cressman, G., 1959: An observational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374.

de Arellano, J. V. G., O. S. Vellinga, and A. A. M. Holtlag, F. C. Bosveld, and H. K. Baltink, 2001: Observational evaluation of PBL parameterizations modeled by MM5. *Proc. 11th PSU/NCAR Mesoscale Model User's Workshop,* Boulder, CO, NCAR, 102–104.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

Gerard, A., and S. Listemaa, 2002: Mesoscale modeling of significant severe weather and flash flood events at WFO Jackson, MS. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., JP4.2.

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

——, J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Tech. Note, NCAR/TN-397+IA, 114 pp.

Janjic, Z. A., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.,* **118,** 1429–1443.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* **47,** 2784–2802.

Manning, K. W., and C. A. Davis, 1997: Verification and sensitivity experiments for the WISP94 MM5 forecasts. *Wea. Forecasting,* **12,** 719–735.

Marzban, C., 1998: Scalar measures of performance in rare-event situations. *Wea. Forecasting,* **13,** 753–763.

Mass, C. F., and Y.-H. Kuo, 1998: Regional real-time numerical weather prediction: Current status and future potential. *Bull. Amer. Meteor. Soc.,* **79,** 253–263.

Pielke, R. A., and Coauthors, 1992: A comprehensive meteorological modeling system— RAMS. *Meteor. Atmos. Phys.,* **49,** 69–91.

Reisner, J., R. J. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.,* **124B,** 1071–1107.

Wang, J., M. R. Hjelmfelt, W. J. Capehart, and R. D. Farley, 2002: Numerical forecast simulations of precipitation events in complex terrain. Preprints, *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*. San Antonio, TX, Amer. Meteor. Soc., JP4.1

Wilks, D., 1995: *Statistical Methods in Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

Zhong, S., and J. Fast, 2003: An evaluation of the MM5, RAMS, and Meso-Eta models at subkilometer resolution using VTMX field campaign data in the Salt Lake Valley. *Mon. Wea. Rev.,* **131,** 1301–1322.