Statistical Characteristics of a Real-Time Precipitation Forecasting Model

BRIAN GAUDET AND WILLIAM R. COTTON

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 20 October 1997, in final form 30 July 1998)

ABSTRACT

At Colorado State University the Regional Atmospheric Modeling System (RAMS) has been used to produce real-time forecasts of precipitation for the Colorado mountain region since 1991. Originally a so-called dump-bucket scheme was used to generate precipitation, but starting in the fall of 1995 real-time forecasts used the bulk microphysics scheme available with RAMS.

For the month of April 1995, a series of 24-h accumulated precipitation forecasts for the month were generated with both the dump-bucket and microphysics versions of the forecast model. Both sets of output were compared to a set of 167 community-based station reports and another set of 32 snow telemetry (SNOTEL) automatic pillow-sensor stations.

The addition of microphysics improved the forecasting of the areal extent and maximum amount of precipitation, especially when compared to the SNOTEL observational set, which is found at locations more representative of the model topography. Climatological station precipitation forecasts were improved on the average by correcting for the difference between a station's actual elevation and the cell-averaged topography used by the model. The model had more problems with the precise timing and geographical location of the precipitation features, probably due in part to the influence of other model physics, the failure of the model to resolve adequately wintertime convection events, and inadequate initializations.

1. Introduction

Accurate quantitative precipitation forecasting (QPF) is one of the most desired aspects of weather prediction to the general community. In addition to the knowledge of whether or not one should go skiing, accurate precipitation forecasts would be of great use to hydrologists and to farmers, who in regions such as Colorado depend on the accumulation of snow in the higher terrain to provide the stream and river water for the growing season. Unfortunately, precipitation is notorious for being difficult to quantitatively predict accurately. The numerical modeler is faced with the problem of predicting a physical process that could be sensitive to any of a number of factors, such as wind, temperature, and humidity, in highly nonlinear ways, including some ways that are not completely understood at the present, such as secondary ice production (Rangno and Hobbs 1994). Another important factor is introduced in regions with highly variable physiography (surface features such as topography, land–water boundaries, vegetation, and soil moisture) on small scales, which can have a profound influence on any of the factors mentioned above. The presence of complex terrain makes even the subsequent determination of the actual precipitation field difficult. Station reports of precipitation in mountain areas are usually sparsely distributed, vary greatly over a distance of only a few kilometers, and are often nonrepresentative of the surrounding terrain. For example, community-based reports of snow amounts in mountainous areas would be expected to be biased toward the amount of snow received in the valleys, since that is the location of most of the towns in mountainous regions. So a direct comparison between model output for a grid cell and that of a town within the grid cell might show a misleading model overestimation of the amount of snow expected in the town.

Regions such as the Colorado Rockies have topography variable on scales finer than is resolvable by existing numerical prediction–assimilation systems such as the Eta Model (Rogers et al. 1996) and the NGM (Hoke et al. 1989). Thus the idea of using a regional mesoscale forecasting model was developed. These models are dependent on the large-scale models for atmospheric initial and boundary conditions but are run at finer grid spacings and have access to fine-scale physiographic information. Certain fine-scale simulations in regions of steep topography have had success in reproducing mesoscale atmospheric features (e.g., Colle and Mass 1996). One would ideally like to use as small a grid spacing as possible for a forecast model, but this

Corresponding author address: Brian Gaudet, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.
E-mail: gaudet@chubasco.atmos.colostate.edu

can increase the computational time to the point at which it no longer becomes possible to produce a real-time forecast. So a balance must be made between grid resolution and forecast time for a given model complexity and computational power.

At Colorado State University the Regional Atmospheric Modeling System (RAMS) model has been used at 16-km fine grid spacing to produce experimental real-time QPF products in winter-type events in the Rocky Mountains of Colorado, using an IBM RISC-6000 workstation that is readily available to the university environment.

This study compares the QPF performance of the real-time forecasting model to two sets of observing stations and to a previous version of the model with more primitive microphysics.

2. Background

As previously mentioned, the demonstration of quantitative skill in numerical precipitation forecasting is still notoriously difficult despite the continuing refinement of atmospheric models. One study of the widely used Nested Grid Model (NGM), a large-scale model used by the former National Meteorological Center [now known as the National Center for Environmental Prediction (NCEP)], examined the model's QPF using the bias score, which measures the ratio of the areal coverage of predicted precipitation to observed precipitation. The study showed that over a 21-month period and a 60-station network the NGM forecasts had a bias slightly greater than one at low precipitation thresholds but closer to 0.5 at higher thresholds (Junker et al. 1989). A more recent study of the NGM, which also evaluated the performance of the Medium-Range Forecast model, found that precipiation forecasting skill varied considerably with season and geography, with smooth topography a major contributor to low forecasting skill (Junker et al. 1992).

More recently, a study by Gartner et al. (1996) found that over the northern plains and Rockies, the 48-km Eta Model had a bias greater than one for precipitation thresholds less than 0.75 in., but that the 29-km Meso Eta Model had a less pronounced bias. Manikin et al. (1996) show that the Meso Eta also outperforms a coarser-resolution version of the Eta (80 km) in another type of skill score, the equitable threat score. However, the improvement was significantly greater in the winter months than in the summer.

An example of regional model quantitative precipitation verification is the study by Anthes et al. (1989), which examined forecasts by the Mesoscale Model, version 4, developed at The Pennsylvania State University and the National Center for Atmospheric Research (the PSU–NCAR MM4 model). For a version of MM4 with 80-km grid spacing, the threat score (see the appendix for definition) showed better model performance at the lower thresholds (scores averaging 0.62 yr^{-1} at the 0.01-in. threshold but 0.35 at the 1.0-in. level, both for 24-h forecasts) and better performance in the winter than the summer. Then a microphysics scheme was added to the model, which explicitly predicted the amount of cloud and rainwater at grid points. The new scheme actually produced lower threat scores, especially at the lower precipitation thresholds (average scores of 0.45 at 0.01 in. and 0.32 at 1.0 in.). The nonmicrophysics model overpredicted the areal coverage of precipitation, but the microphysics model underpredicted the area at all but the largest threshold. This was attributed to the failure of the model to achieve saturation over a grid cell often enough, due to the coarseness of the model grid spacing. The simulations were performed over the eastern United States, and the model output was verified against an objective analysis of station data.

A follow-up study by Giorgi and Bates (1989) used a 30-day simulation of the MM4 over the western United States to see if their model could reproduce the climatology of precipitation there. They found that both the MM4 and a version with more accurate surface layer and radiation schemes produced similar threat scores, slightly lower than those of Anthes et al. (around 0.38 at the 0.01-in. level). This was attributed to, among other reasons, the complex topography found in the western United States. Rather than objectively analyzing a precipitation dataset, and comparing it to model grid output, they interpolated the model grid output to the locations of the stations, similar to the procedure that we will use (see section 5).

The NGM and MM4 studies serve to emphasize that the use of more sophisticated physics in a computer model does not guarantee improvement in precipitation forecasting if the resolution of the model (or the data used by the model) is not sufficient—this finding will also be shown by our research.

3. QPF using the RAMS model

At Colorado State University, RAMS has been used for real-time forecasting since 1991 (Cotton et al. 1994). RAMS is a nonhydrostatic primitive equation model developed at Colorado State in the early 1980s (Tripoli and Cotton 1982; Pielke et al. 1992) and is capable of simulating phenomena from the continental scale to the scale of turbulent eddies. Standard versions of RAMS predict the u, v, and w components of the wind; the Exner function [$\Pi = c_p(p/p_0)$, where c_p is the dry specific heat of air at constant pressure, p is the pressure, and p_0 is a reference pressure at sea level]; dry air density ρ; ice–liquid potential temperature θ_{il} (Tripoli and Cotton 1981); and the mixing ratios of various forms of water, described below. These variables are predicted on one grid or a series of nested grids, each with two-way interaction across the interfaces (Clark and Farley 1984). The grid coordinates are based on a polar stereographic projection in the horizontal directions and on a terrain-following (σ_z) coordinate in the vertical (Gal-

Chen and Sommerville 1975). An Arakawa C-grid concept is used, where the variables u, v, and w are defined at locations staggered one-half grid length in the x, y, and z directions, respectively, from the thermodynamic and pressure variables (Arakawa and Lamb 1981). Terms relating to the propagation of acoustic waves are integrated with a time step shorter by a factor of 3 in order to maintain computational stability (Klemp and Wilhelmson 1978).

At first, the forecast model version of RAMS used a so-called dump-bucket parameterization to represent winter precipitation, which is a crude parameterization similar in concept to that developed by Rhea (1978). In our scheme, the highest saturated layer in regions of upward motion is found (about 96% for our grid spacing). A precipitation efficiency is then calculated; it is a function of the temperature of this layer, specifically

$$E = \min(-0.01T_{\text{sat}}, 0.25), \tag{1}$$

in which E is the precipitation efficiency and T_{sat} is the temperature in degrees Celsius of the highest saturated level. This efficiency gives the fraction of the total liquid water content above saturation that is converted into precipitation.

This scheme has the advantage of being very quick computationally. It suffers the disadvantages of 1) not including any of the microphysics of the precipitation-producing process, and 2) not including any interaction of precipitation with its environment as it falls. Thus no direct evaporative or condensational effects on hydrometeors, as well as the different behaviors of the various species, are included.

Additionally, the dump-bucket version of the forecast model used a modified Kuo cumulus parameterization scheme (Tremback 1990) in an attempt to simulate convective precipitation, which is not uncommon in the winter season over mountainous terrain. The parameterization did contribute somewhat to the total month's precipitation, but overall the contribution of the convective parameterization to the forecast accuracy was negligible. The scheme was not designed for use with winter season convection, nor is such a scheme generally used for grid spacings as small as 16 km.

The microphysics used in the current version of the forecast model is that of version 3b of RAMS and is described in numerous sources (Walko et al. 1995). It is a bulk microphysics scheme, meaning that the total mixing ratio over all diameters for a particular water species is predicted but not the amount of species at any specific diameter. However, in RAMS the user specifies that the size spectra of water species over diameter follow a particular gamma distribution, and this provides the information needed to determine the mass of the species at a particular size. The generalized gamma distribution has the form (Flatau et al. 1989; Walko et al. 1995)

$$f_{\text{gam}}(D) = \frac{1}{\Gamma(\nu)}\left(\frac{D}{D_n}\right)^{\nu-1}\frac{1}{D_n}\exp\left(-\frac{D}{D_n}\right), \tag{2}$$

where D is the particle diameter, ν is the shape parameter that varies the breadth of the distribution, and D_n is a characteristic diameter defined as D_{mean}/ν. The model predicts the mixing ratios of six condensed water species at all grid points: rain, pristine ice, snow (essentially pristine ice large enough for significant riming), aggregates, graupel, and hail. Equations governing the growth, evaporation–sublimation, and interconversion of these species are used to prognose their mixing ratios, as well as advection and subgrid-scale diffusion. The total mixing ratio of water is also a predicted variable; whatever total water is not accounted for in the six species listed above is assumed to be water vapor and cloud water. The cloud water portion is any amount exceeding the saturation mixing ratio (thus no supersaturations are maintained in the model), and water vapor is any water substance left over. Though the mixing ratio of cloud water is not directly predicted, the number concentration of cloud droplets is specified (300 cm^{-3} for these forecasts). The number concentration of pristine ice is also a predicted variable, using the ice nucleation equations described in Walko et al. (1995). These empirically based equations include contact freezing, the homogeneous freezing of cloud drops and haze particles, and direct nucleation from the vapor; this last process is expressed as a function of ice supersaturation (Meyers et al. 1992).

Past case studies of microphysics suggested that the inclusion of microphysics in the real-time model greatly enhanced forecasting skill, compared with a dump-bucket-based model (Thompson 1993; Beitler 1994; Cotton et al. 1994; Cotton et al. 1995). Generally, the dump-bucket model had a tendency to underestimate the total amount of precipitation by large amounts in events with strong synoptic forcings, which was corrected by the inclusion of microphysics.

4. Current study

The forecast model used for this study used two grids: a 39×31 coarse grid with 80-km grid spacing covering the western United States, and a 42×37 fine grid with 16-km grid spacing centered over Colorado (see Fig. 1). The eastern third of the model domain was mostly plains, while mountains and valleys dominated the western two-thirds of the grid (Fig. 2). The time steps used were 90 s on the coarse grid and 45 s on the fine grid. The model uses 26 vertical levels, with grid spacing gradually increased from 250 m near the surface to 1 km near the model top at 17 km.

The vegetation type (and corresponding surface parameters) was allowed to vary across the model domain and was taken from the Biosphere–Atmosphere Transfer Scheme database of 18 vegetation types (Dickinson et al. 1986). Geographical datasets of topography (at 30-

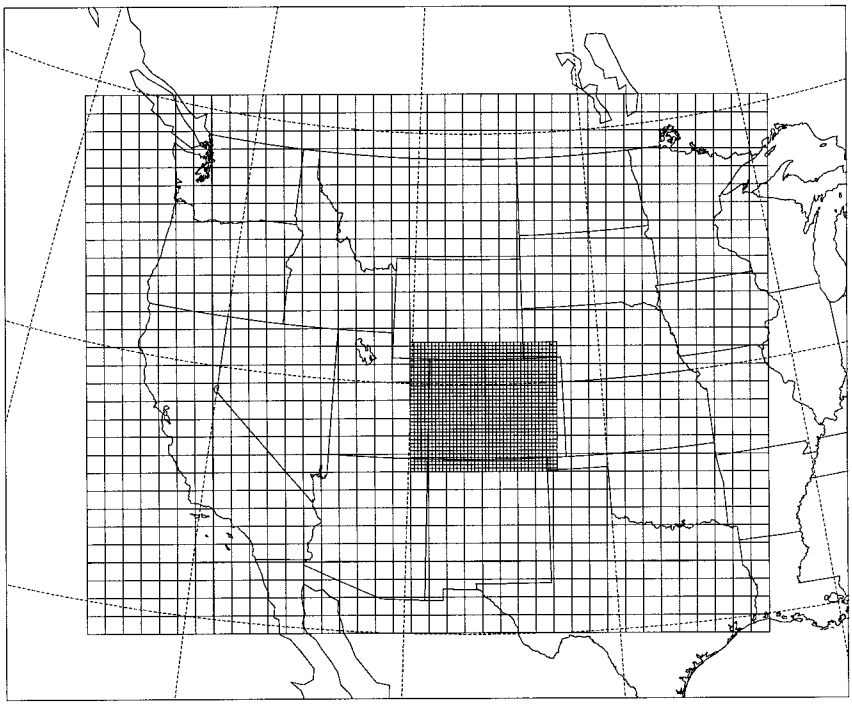

FIG. 1. Location of forecast model grids.

s increments), percentage of surface water, and month-dependent sea surface temperatures were also available for use. Two schemes for converting the topography data to elevation at a grid point are available to RAMS. The first is essentially averaging all the dataset elevation values within a grid cell and using the average as the model elevation.[1] The second scheme is the silhouette average, in which the maximum dataset elevation for each cross section through the grid cell is taken as the silhouette height, and then the silhouette heights for all the cross sections are averaged. The first scheme has the disadvantage of reducing topographic relief (Jarraud et al. 1987), but the second scheme makes it more difficult to better resolve topography with decreased grid spacing. For its topography representation, the forecast model uses a weighting of 75% for the silhouette average and 25% for the conventional average.

Until April of 1995 the winter season had been abnormally dry in Colorado, raising fears that the summer would be characterized by drought conditions. However, during April 12 storm systems crossed the state (according to N. Doesken in the April 1995 weather summary for the Colorado *Climatological Data* publication). The Arkansas and Platte drainage basins received twice their normal monthly precipitation, while the Colorado and Rio Grande basins received 50% more than average. After an equally wet May, the snowpack for the state surged above normal and resulted in some flooding toward the end of the month as the snowpack began to melt. The frequency of precipitation events (mostly of the winter stratiform precipitation type) made this a good month to compile statistics to evaluate the models' forecast skill. During the month of April 1995, the forecast model was based on a version of the model with the dump-bucket precipitation scheme substituted for the model microphysics; for more details refer to Cotton et al. (1994).

The forecast initialization was usually created from the 0000 UTC Rapid Update Cycle (RUC) analyses generated by NCEP. These analyses possess 60-km grid spacing and are based on the hybrid-b Mesoscale Analysis and Prediction System framework, which is described in Bleck and Benjamin (1993). To supplement the RUC data, synoptic surface observations and rawinsonde information were incorporated using a Barnes objective analysis scheme (Barnes 1973). Usually the RUC files were available by 0400 UTC, and so the RAMS forecast cycle would commence at 0430 UTC. The forecast model also requires lateral boundary conditions at future times out to 36 h, and these were provided by the 80-km Eta forecasts available at the time (Mesinger et al. 1988). A ''nudging'' procedure was followed, where the Eta 12-, 18-, 24-, 30-, and 36-h forecasts were used to create tendencies to the RAMS model fields that would tend to converge them to the Eta solution at grid points near the coarse grid bound-

[1] A $4\Delta x$ filter is applied to the topography during both topography averaging processes.

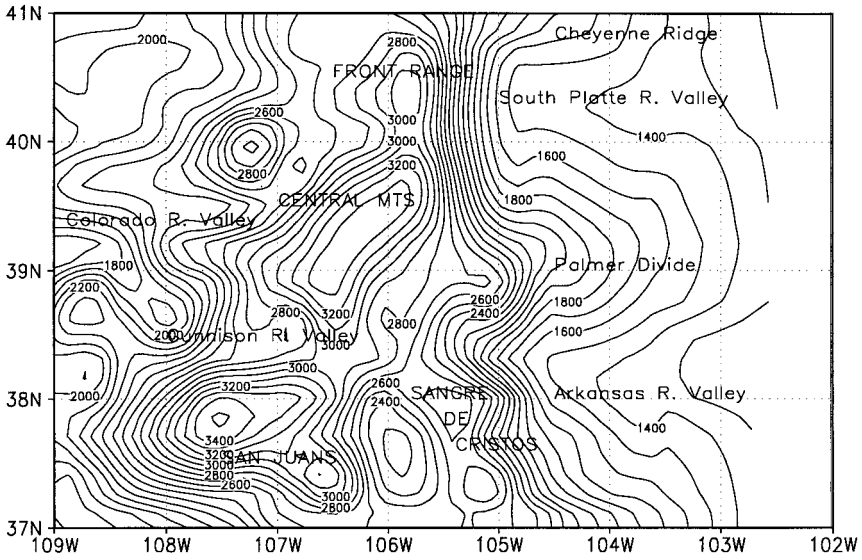

FIG. 2. Elevation of the terrain on grid 2, in m. Labels in capital letters mark the location of mountain ranges, while mixed letters denote rivers or drainaige divides.

aries. Note that this does not mean that the RAMS solutions near the boundaries and the Eta forecasts were identical, nor did the nudging directly influence the RAMS fine grid, except by advection from the coarse grid. The 80-km Eta analysis for 0000 UTC was used for initialization on one day when the RUC analysis was unavailable. The model run normally reached the 36-h point by 1100 UTC (0500 LT), at which time the output was sent through a graphics package, and the resultant plots were displayed on the Internet.

Among the model output was the total accumulated precipitation since the start of the simulation on the fine grid, stored at 2-h increments. Thus, for a particular observing station, the amount of model forecast precipitation at that site during its verification period could be reconstructed with 2-h resolution.

In late 1995, the IBM RISC-6000 370 that had been used to perform the real-time forecasts was replaced by a 390-series computer, which possesses approximately twice the computing speed. At about the same time, a new version of RAMS was introduced, along with its optimized microphysics (Walko et al. 1995). So the real-time forecasts began incorporating RAMS bulk microphysics. The days in the month of April 1995 were then resimulated using the new version of the forecast model. Only model precipitation generated in the first 24 h of the real-time forecasts were used for statistical purposes, because the new microphysics simulations were only carried out for 24 h. Some days of the month were also resimulated with the dump-bucket version of the forecast model, because on those days a real-time forecast was not produced due to failures in the forecast cycle.

5. Observational data

Two sets of observational data were used. One consisted of the set of stations reporting data to the Colorado edition of the National Climatic Data Center (NCDC 1995) publication of daily and monthly meteorological data, minus those stations either not within the RAMS fine grid or not possessing a complete record of precipitation for April 1995.[2] These stations will be referred to as climate stations, and there were 167 of them. Precipitation at these stations are measured on a daily basis by rain gauges after any solid precipitation is melted. Human observers perform the measurement at all of these stations except for Denver International Airport, which uses an automated system. The set of climate stations shows good coverage over the state of Colorado (see Fig. 3). But most of these stations are located in communities and, as alluded to previously, will not be representative of precipitation amounts at the highest terrain in the mountainous regions.

The second set of data consisted of snow telemetry (SNOTEL) stations. These are automated pillow-sensors located in remote mountainous regions throughout the western United States and are maintained by the U.S. Department of Agriculture's Natural Resources Conservation Service (McMillan 1981). Data include daily maximum–minimum temperatures, mean daily temperatures, and the total water-equivalent precipitation since the start of the water year, 1 October. The data are transmitted by radio waves from the stations and were obtained from a Western Regional Climatic Center archive at the Desert Research Institute in Nevada. The data are transmitted daily or sometimes twice daily, but gaps exist in the data record at all stations. No stations transmitted data on 12–13 April, nor from 19–20 April dur-

[2] Actually, the first and last days of the month were excluded, leaving 2–29 April as the observational period.

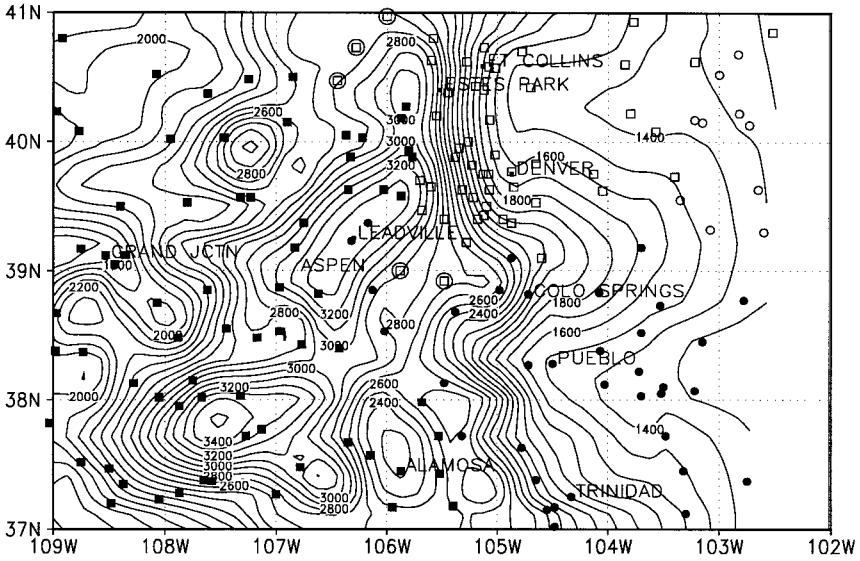

FIG. 3. Location of climatological data stations used for study. Stations in the Platte, Republican, and Arkansas drainage basins are denoted by open squares, open circles, and closed circles, respectively. Closed squares denote stations in the Colorado and Rio Grande basins. Open squares that are encircled indicate, from north to south, Hohnholz Ranch, Walden, Spicer, Antero Reservoir, and Lake George.

ing one of the snow episodes. A total of 32 Colorado SNOTELs transmitted data during the 25 other days from 2 to 29 April, and these stations formed the dataset (Fig. 4). The SNOTELs generally receive more precipitation than neighboring climate stations (Doesken and Schaefer 1987). They are more representative of the snowpack accumulation locations than the climate stations and so should provide a better basis for estimating model performance in those regions.

One complicating factor to these sets of observational data is that they do not all verify at 0000 UTC. The SNOTEL stations all report data at midnight local standard time, which is 7 h after the beginning of the model runs. So, to compare the forecast model amounts with the observations, precipitation the last 18 h of a model run was combined with that of the first 6 h of the next day's run. The procedure becomes even more complicated for the climate stations because different climate stations report data at different times (except for the synoptic airways stations, which always verify at midnight). For each climate station, the verification time was used to reconstruct how much model precipitation

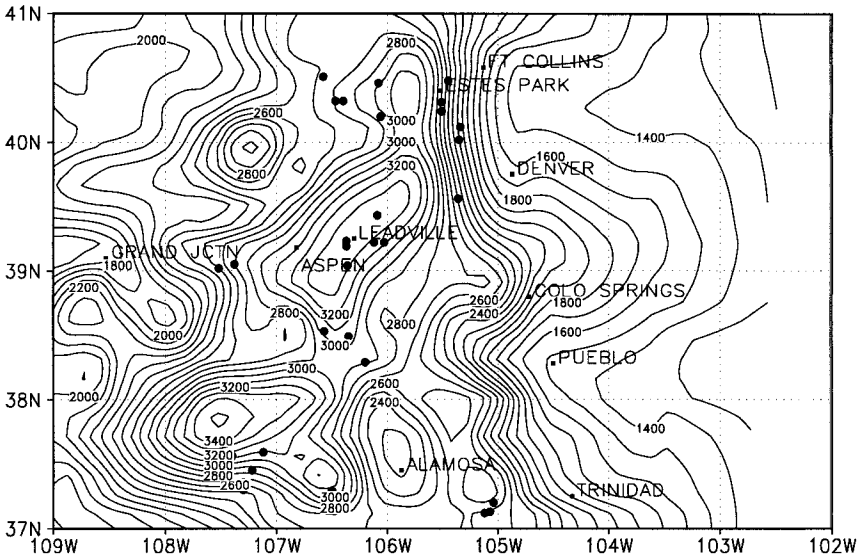

FIG. 4. Location of SNOTEL data stations used for study.

was produced for its verification period. A consequence of this procedure, though, is that the period of 9 April for one climate station does not exactly correspond with the 9 April period for other climate stations.

Another potential problem is the fact that the traditional 8-in. rain gauges used at the climate stations often underestimate frozen precipitation, particularly in windy conditions, because the precipitation drifts away from the containment vessel (Doesken and Judson 1997). The undercatch can exceed 50% in blizzard conditions (Goodison 1978). The SNOTEL pillow-sensors are better at retaining the snow that falls on a given area. Therefore, a climate station can report considerably less precipitation than a SNOTEL even if they are in close proximity (Doesken and Schaefer 1987).

A third potential area of concern is that at the beginning of a simulation precipitation amounts may be greatly in error until the model has a chance to ''spin up;'' usually, during this period little precipitation is generated. Such a spinup time was found in the monthly statistics of the current study but only for about 2–4 h after initialization, which is consistent with the findings of previous nonhydrostatic RAMS studies (Snook and Pielke 1995). We assume that the spinup time does not signficantly change the monthly statistical trends found; indeed, it should help counteract the gauge undercatch effect, though it imposes a further 10%–20% uncertainty in the results. Our spinup underprediction seems mainly due to cases in which the model initialization does not properly include all the mesoscale dynamics present in the atmosphere.

One final fact to keep in mind is that the observations are all point measurements of precipitation, and we will usually take the approach of comparing these data with the precipitation the model produces at the grid point nearest to the station. An attempt to compare these observations to model output, which is essentially an average over a grid volume, will produce certain inherent errors. It could be argued that for a true comparison one would be forced to smooth the observational precipitation field and compare this with the model output in a qualitative sense. But quantitative measurements based on this procedure can be misleading if a localized feature causes an extreme station data point, since the objective analysis would spread the influence of this data point to regions where there is simply no observational data. So when there is reason to believe that the observational data field may be highly variable, it may be safest to use only ''actual'' observed data and rely on the resolution of the model to reproduce small-scale features.

6. Quantitative results

Though much can be learned from a simple qualitative analysis of model forecast features, it is difficult to draw conclusions of any certainty without the use of quantitative statistical procedures. In this application these mainly consist of finding some ''average error'' between model output and observations, and determining characteristics of this error. Probably no one statistic is completely satisfactory in expressing skill, so several measures of our models' skill will be examined here.

a. Total model precipitation biases

Perhaps the easiest statistic from this analysis to understand is the total amount of precipitation observed at all the stations divided into the precipitation produced by the forecast model at the set of grid points closest to each observational station, $p_{\text{mdl}}/p_{\text{obs}}$. This will be referred to as the total precipitation bias and gives an indication of general overprediction or underprediction by the models but not how the total precipitation is distributed among the stations or the observing period. The dump-bucket model had a bias of 0.87 for the climate stations and 0.38 for the SNOTELs; for the microphysics model, the biases were 1.33 and 0.78, respectively. The dump-bucket model is the drier of the two models for both station sets. For the set of climate stations, the dump-bucket model is slightly closer to predicting the true amount of total precipitation but predicts much too little precipitation compared to what is actually observed at the SNOTELs.

One can also find the precipitation model biases of each station and then create a geographical plot of precipitation ratio for each of the forecast models. Figure 5 shows the distribution of the dump-bucket model biases for the climate station set, after being objectively analyzed. The plot generally shows underprediction of precipitation except in various river valleys (the Gunnison and Colorado in the west; the Platte and Arkansas in the east) and in places in the San Juan Mountains. There is major underprediction in the northwest, the Palmer Divide region, and the central mountains. The plot seems to show that the dump-bucket model has problems in representing the increase of precipitation with elevation, especially in the central mountains.

In contrast, the only large regions where the microphysics model (Fig. 6) underpredicts precipitation at the climate stations are the Arkansas River valley and a number of valleys in the west. Large overpredictions occur in the San Juan Mountains and parts of the northern Rockies, and are most extreme near the steepest terrain.

Similar plots can be created using the SNOTEL stations, but these sites show a small areal coverage over the state (the stations are mostly concentrated in the San Juan Mountains and the northern Front Range), so Fig. 7 simply plots the bias at each station. It indicates that the microphysics model slightly overpredicts the amount of precipitation in parts of the slopes of the Front Range and San Juan Mountains, but not by as much as the climate station record would indicate, which could be due to the undercatch problems of climate stations. Precipitation is somewhat underpredicted at SNOTELs

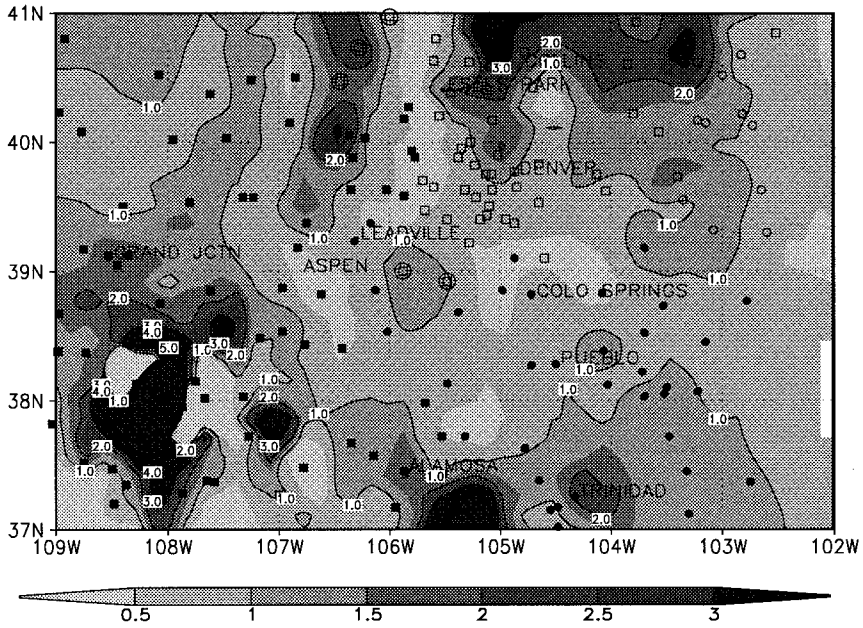

FIG. 5. Ratio of total model precipitation to total observed precipitation using dump-bucket scheme, evaluated at climatological stations. Model precipitation for a station is taken to be the precipitation produced at the grid point nearest to the station. Data has been subjected to a Cressman objective analysis (Cressman 1959).

near the crest of the previously mentioned mountain ranges, which suggests that inadequate resolution of the topography could also be a factor.

b. Scores based on set theory

Precipitation has traditionally been verified by organizations such as NCEP with bias scores and threat scores (Anthes 1983; Junker et al. 1989; Schultz 1995). The bias score *B*, mentioned in section 2, is the ratio of the number of stations forecasted to reach a certain precipitation threshold to the number of stations that actually reach the threshold; a perfect forecast would have $B = 1$, while values of B less than and greater than one represent underforecasting and overforecasting, respectively, the precipitation areal coverage. (This is in contrast to the bias mentioned in the previous section, which is a ratio between amounts of precipitation, not number of stations receiving precipitation.) The bias score does not measure the coincidence of stations that have precipitation forecasted and those at which it is observed, whereas the threat score T is a measure of the correlation between forecasted and observed stations. These scores are defined in the appendix, as is the Heidke skill score (HSS), which is similar to the threat score but has a couple of beneficial features (it takes into account correctly forecasting the nonoccurrence of precipitation and subtracts out the skill level expected purely from chance).

Tables 1 through 3 show a comparison between the microphysics and dump-bucket model performances with the above-mentioned skill scores for the climate stations. Here is a summary of some of the trends indicated by the tables.

1) At low precipitation thresholds, both models show precipitation biases greater than one, but at the higher thresholds the microphysics model still exhibits a bias score greater than one on average, whereas the dump-bucket model shows a bias score less than one. Of the two models, the microphysics shows biases closer to one at the lowest threshold. The larger biases of the dump-bucket model at this amount could be due in part to the cumulus parameterization. The relative performance of the dump-bucket model in the bias score is best at the 2.54-mm (0.1 in.) level, while the microphysics model again shows the better bias scores at the highest thresholds. (It should be noted that the sample size is small at the high thresholds, so there is a larger potential variability in skill scores at these levels.)

2) The threat scores seem to show no significant difference in performance for either model at all except the highest (12.7 mm) threshold.

3) The Heidke skill scores show better performance for the microphysics model at all thresholds [though the difference is slight at the 6.35-mm (0.25 in.) level] despite the similar threat scores. At the lower thresholds, generally the dump-bucket model forecasts have larger bias scores. As a consequence, this model's forecasts have a higher expected threat score due solely to chance, so a given threat score implies less true skill for the dump-bucket model forecasts. At the 12.7-mm (0.5 in.)

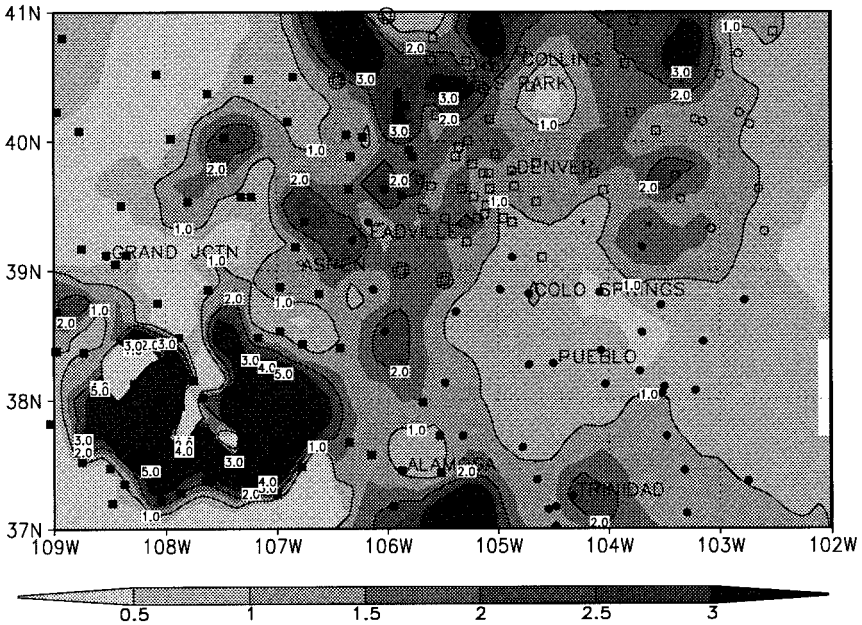

FIG. 6. Same as Fig. 5 but using microphysics model.

threshold, the dump-bucket version of the forecast model simply did not generate that amount of precipitation often enough. For both models, the highest Heidke skill score (as well as the best bias scores) occurred at the 2.54-mm threshold.

In summary, the microphysics model seems to be better than the dump-bucket model at reproducing the spatial and quantitative extent of precipitation. Specifically, the dump-bucket model is more prone to produce light precipitation where in fact no precipitation was observed and usually underestimates the maximum amount of precipitation expected from an event. So the use of the microphysics model should improve yes–no forecasts of precipitation and be more useful in predicting accumulated precipitation at stations in at least a climatological sense, which is similar to the findings of Giorgi and Bates (1989). The microphysics model can have trouble in predicting the proper geographic location of precipitation features for a specific event, however, often due to the timing of frontal features. Quantitatively, our threat scores are considerably lower than those of the Anthes et al. (1989) study, performed over the eastern United States, but higher than the scores of the Giorgi and Bates study, which like ours were for stations in the western United States. This reiterates the forecasting difficulties endemic to mountainous regions.

7. Systematic errors

Model precipitation forecasts, of course, can vary considerably from the observations if various processes or finescale features are present in nature but not in the model. Or, it is possible that precipitation may be reproduced in the model but not captured with the observational method. Ultimately, the best that one can hope for with such errors is that they are random in nature and do not introduce major systematic biases to the forecasts. However, this section will identify biases that could be responsible for some of the trends previously mentioned.

a. Model versus actual elevation

The elevation a model uses for a grid point is normally not the elevation of the corresponding point in geographical space to the resolution of available data but is obtained by averaging the actual elevations over a region corresponding to a grid cell surrounding that point. If there is considerable variation in actual topography within that grid cell, it can be expected that observed precipitation within the cell could differ drastically from what the model produces. Interpolation can alleviate some of this problem but not when the observing station is surrounded by terrain thousands of meters higher.

If the grid cells are sufficiently small, it can be hypothesized that the dominant influence on observed snowfall is the station elevation, since on that scale the horizontal atmospheric variability is small. So we can represent the precipitation distribution within a given grid cell as $p = p(z)$, where z is the elevation. Note that the particular form of $p(z)$ need not be the same for different grid cells (certainly not if they are far apart geographically), but one would generally expect that p is an increasing function of z.

Now let us further assume that the forecast model is

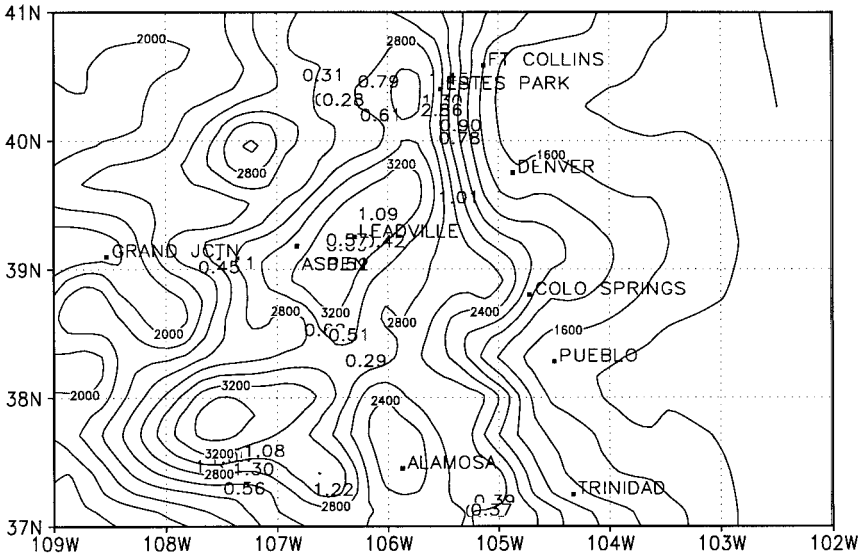

FIG. 7. Station plot (large labels) of ratio of total microphysics model precipitation to total observed precipitation for SNOTEL stations. Topography contour interval is 200 m.

ideal in the sense that it produces the observed amount of precipitation at a grid point if the gridpoint elevation Z is the same as the actual elevation at that point, z_0. So, if we denote the actual elevation of a station as z_s, we can represent the actual precipitation at any other point in the grid cell as a Taylor series:

$$p(z) = p(z_s) + \frac{dp}{dz}(z - z_s) + \frac{1}{2}\frac{d^2p}{dz^2}(z - z_s)^2 + \ldots, \tag{3}$$

where $p(z_s)$ is the actual amount of precipitation observed at the station if we make the above idealization hypothesis, z is the elevation of another point in the grid cell, and all derivatives are evaluated at the station elevation z_s. The precipitation produced by the model at the nearest grid point to the station would then be $p(Z)$. If this equation is valid, with the most important derivative of p being the positive value dp/dz, then we would have overprediction of precipitation at stations whose elevation is less than that of the nearest model grid point, if we use that point's precipitation to represent the station.

b. Experimental verification

For the month simulated for this study, the forecast model significantly overpredicted the precipitation at the climate stations (by 33%), whereas for the SNOTEL

TABLE 1. Skill scores for dump-bucket model at the 167 climate stations. Labels on heading denote thresholds used to evaluate each score, in mm of precipitation. Only days with significant observed precipitation are shown.

	Bias score				Threat score				Heidke skill score			
Date	0.254	2.54	6.35	12.7	0.254	2.54	6.35	12.7	0.254	2.54	6.35	12.7
8 Apr	7.00	3.00	0.00	0.00	0.07	0.00	0.00	0.00	0.05	−0.05	0.00	0.00
9 Apr	1.43	0.84	0.36	0.11	0.55	0.26	0.20	0.00	0.00	0.00	0.22	−0.01
10 Apr	1.30	1.22	1.34	0.88	0.77	0.71	0.55	0.39	0.00	0.42	0.44	0.43
11 Apr	1.54	1.47	1.36	0.00	0.57	0.36	0.13	0.00	0.37	0.32	0.17	0.00
16 Apr	2.12	0.74	0.86	0.00	0.44	0.17	0.18	0.00	0.19	0.14	0.28	0.00
17 Apr	1.65	1.74	1.77	1.00	0.57	0.45	0.32	0.11	0.07	0.30	0.33	0.15
18 Apr	1.16	1.09	0.84	0.39	0.80	0.55	0.29	0.14	0.15	0.29	0.23	0.19
19 Apr	1.29	1.07	0.80	0.09	0.77	0.58	0.29	0.00	0.14	0.42	0.26	−0.02
20 Apr	1.49	1.42	1.58	0.00	0.65	0.50	0.17	0.00	−0.02	0.36	0.14	0.00
21 Apr	1.49	1.09	0.42	0.27	0.66	0.45	0.15	0.08	0.03	0.36	0.16	0.12
22 Apr	1.51	1.26	0.43	0.35	0.66	0.45	0.18	0.15	0.16	0.22	0.16	0.22
23 Apr	1.65	0.47	0.43	0.00	0.55	0.17	0.05	0.00	0.22	0.15	0.05	0.00
24 Apr	1.57	1.16	0.43	0.25	0.61	0.44	0.18	0.00	0.19	0.35	0.22	−0.01
25 Apr	2.00	0.75	0.83	0.00	0.31	0.33	0.22	0.00	0.32	0.46	0.34	0.00
26 Apr	1.15	0.77	0.38	0.20	0.83	0.59	0.33	0.11	0.53	0.43	0.29	0.13
27 Apr	3.87	0.60	0.00	0.00	0.13	0.14	0.00	0.00	0.02	0.21	0.00	0.00
29 Apr	1.35	0.47	0.13	0.00	0.55	0.18	0.08	0.00	0.17	0.09	0.13	0.00

TABLE 2. Same as Table 1, but for microphysics model.

	Bias score				Threat score				Heidke skill score			
Date	0.254	2.54	6.35	12.7	0.254	2.54	6.35	12.7	0.254	2.54	6.35	12.7
8 Apr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9 Apr	0.80	0.55	0.43	0.78	0.44	0.34	0.15	0.07	0.17	0.27	0.11	0.08
10 Apr	1.27	1.19	1.53	2.05	0.79	0.72	0.59	0.44	0.15	0.46	0.48	0.42
11 Apr	1.62	1.86	4.55	7.67	0.57	0.43	0.17	0.08	0.35	0.40	0.21	0.13
16 Apr	1.20	0.89	0.86	1.00	0.57	0.50	0.08	1.00	0.52	0.58	0.12	1.00
17 Apr	1.19	1.44	2.03	3.80	0.62	0.42	0.27	0.17	0.40	0.28	0.24	0.22
18 Apr	1.00	1.11	1.45	1.94	0.74	0.61	0.30	0.10	0.23	0.40	0.16	0.05
19 Apr	1.05	0.94	1.10	0.91	0.83	0.57	0.41	0.16	0.59	0.44	0.40	0.16
20 Apr	1.35	1.62	2.33	2.75	0.66	0.47	0.14	0.00	0.14	0.27	0.06	−0.08
21 Apr	1.21	0.85	0.70	0.91	0.55	0.31	0.10	0.05	−0.05	0.18	0.02	0.03
22 Apr	1.30	1.28	1.21	1.82	0.58	0.44	0.44	0.23	0.08	0.21	0.44	0.28
23 Apr	0.99	0.98	1.29	4.00	0.56	0.39	0.23	0.00	0.44	0.42	0.31	−0.01
24 Apr	1.47	1.45	1.75	2.50	0.60	0.50	0.31	0.08	0.22	0.41	0.32	0.11
25 Apr	0.56	0.38	0.17	0.00	0.31	0.29	0.17	0.00	0.41	0.42	0.28	0.00
26 Apr	0.88	0.78	0.69	0.88	0.77	0.65	0.55	0.43	0.56	0.54	0.49	0.48
27 Apr	3.61	2.90	5.00	11.00	0.19	0.15	0.04	0.09	0.13	0.18	0.05	0.16
29 Apr	0.54	0.14	0.00	0.00	0.36	0.08	0.08	0.00	0.21	0.07	0.00	0.00

stations total precipitation was underpredicted by a similar amount (22%). How much of this difference can be explained through a model height argument? Figure 8 shows that for the climate stations the elevation at the grid point nearest to a station is almost invariably higher than the actual elevation of the station, which would result in a systematic overprediction of the total precipitation amount by the arguments presented above. The SNOTEL stations do not show this tendency. Thus, at least for the climate stations, the model height argument is consistent with the observations and can be further tested.

This suggests the use of a month's accumulated precipitation data for the climate stations to verify the applicability of (3). The problem is the determination of the functional form of $p(z)$, which can vary from grid cell to grid cell. One approach is to assume that $p(z)$ will have approximately the same form for geographically related regions and then to use the accumulated model output to find an average $p(z)$ over that region. If such a function of z can be found, it is reasonable to determine the derivatives of $p(z)$, and then for each grid cell in the region use (3) to find the true precipitation $p(z_s)$ at each station, using derivatives evaluated at z_s. Spreen (1947) found a precipitation–elevation curve based on observed precipitation in western Colorado for a single season. A single linear relationship between precipitation and elevation was found to be as effective as any other polynomial fit to the data, though the remaining variance was large and could not be explained without taking slope, exposure, and compass direction of exposure into account.

The climate stations in Colorado are classified by the NCDC as belonging to one of five drainage basins: the

TABLE 3. Median skill scores for climate stations. Only days with sufficient precipitation (8–11, 16–27, 29 April) were used in the sample.

Score (mm)	Dump bucket	Microphysics
B, 0.254	1.51	1.19
2.54	1.09	0.98
6.35	0.43	1.21
12.7	0.00	1.82
T, 0.254	0.57	0.57
2.54	0.44	0.43
6.35	0.18	0.17
12.7	0.00	0.09
HSS, 0.254	0.14	0.22
2.54	0.30	0.40
6.35	0.17	0.21
12.7	0.00	0.11

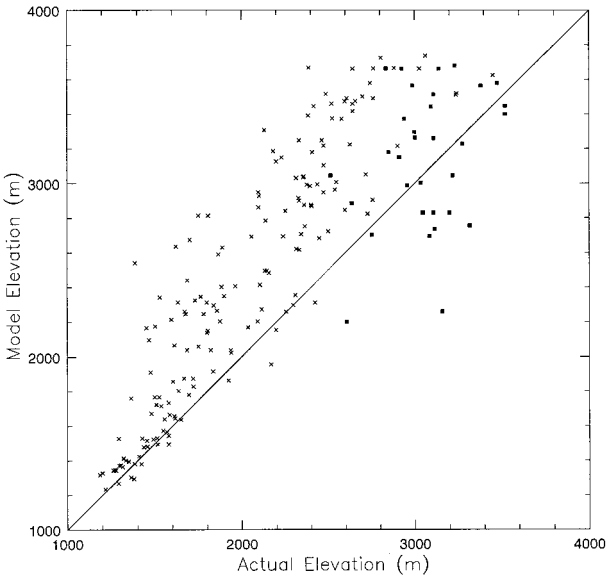

FIG. 8. Model elevation at nearest grid point to station vs station elevation for climatological stations (crosses) and SNOTEL stations (squares).

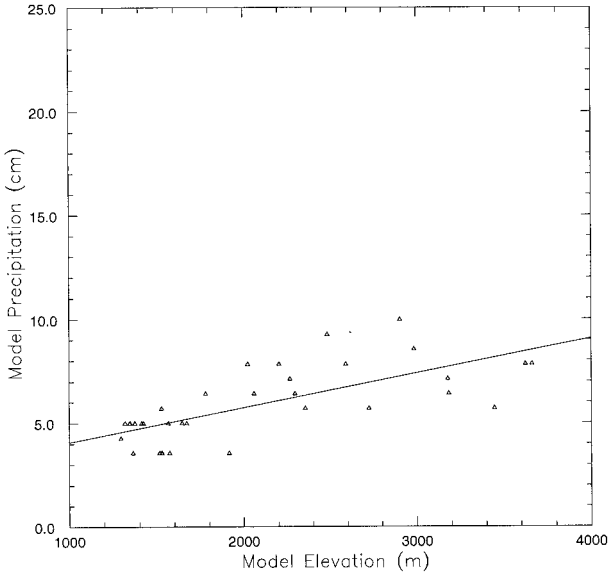

FIG. 9. Total model precipitation generated by the 24-h microphysics simulations of April 1995 as a function of elevation for climatological stations found in the Arkansas drainage basin. Both precipitation and elevation are those of the model grid point nearest to each station. Curve represents the equation $y = 0.001663x + 2.415$.

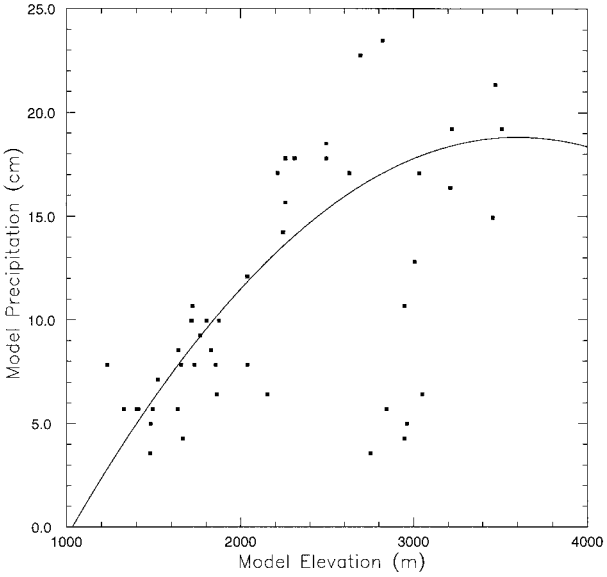

FIG. 10. Same as Fig. 9, but for climatological stations in the Platte drainage basin. Curve represents the equation $y = (-2.843 \times 10^{-6})x^2 + 0.02050x - 18.146$.

Arkansas, the Colorado, the Kansas, the Platte, and the Rio Grande. It was discovered that if the stations were grouped into the categories Arkansas, Platte, and Colorado–Rio Grande, then plots of total model precipitation at the grid points nearest to each station versus model elevation at these grid points display a usable functional form (the Kansas River basin is wholly on the plains of eastern Colorado, and so these stations were not used for this analysis). The relationship between p and z is mostly linear, with the slope in the Arkansas drainage basin less than the slope in the other two regions (see Figs. 9–11). The Platte River basin is mostly on the east side of the Rocky Mountains, and a station at a given elevation in this region received about 10 more centimeters than the corresponding elevation in the Colorado and Rio Grande basins, due to the rain-shadow effect. The exception is five stations in the Platte basin (Lake George, Antero Reservoir, Spicer, Walden, and Hohnholz Ranch), which are located in intermontane valleys west of the Front Range; these stations received much less rainfall than others in the Platte basin at that elevation and fit better with the stations in the Colorado and Rio Grande basins (see Fig. 3 for the locations of these stations). Furthermore, some stations near 3000 m in the Platte basin lie well above the others, suggesting that a further geographical decomposition may be possible.

A linear fit was made for the Arkansas stations, since they showed little variation in precipitation with elevation. However, for the other plots a quadratic term had to be included, to take into account the reduction of dp/dz at high elevations in the Platte basin and to keep the curve from reaching zero at low elevations in the Colorado–Rio Grande basins. These plots were fitted with parabolas that minimized the root-mean-square deviation (with the five stations mentioned above excluded from the Platte basin analysis). In the case of quadratic curves, the last two terms in (3) can be evaluated and then solved for the corrected monthly precipitation $p(z_s)$ using the actual station elevation.

Figure 12 combines the climate station data for all

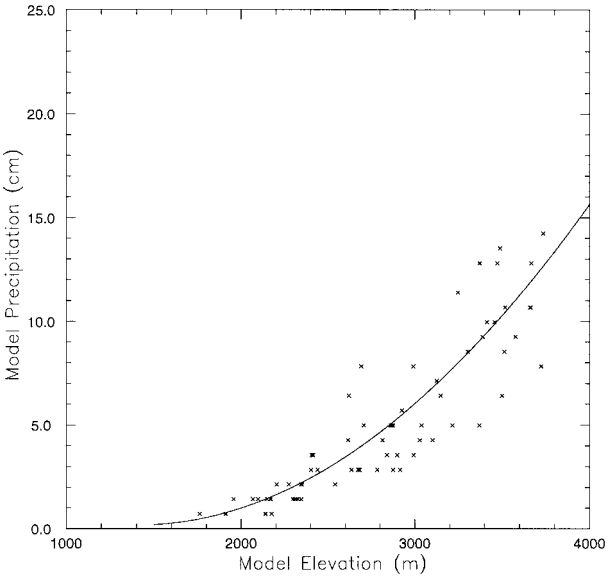

FIG. 11. Same as Fig. 9, but for climatological stations in the Colorado and Rio Grande drainage basins. Curve represents the equation $y = (2.291 \times 10^{-6})x^2 + 0.006423x + 4.677$.

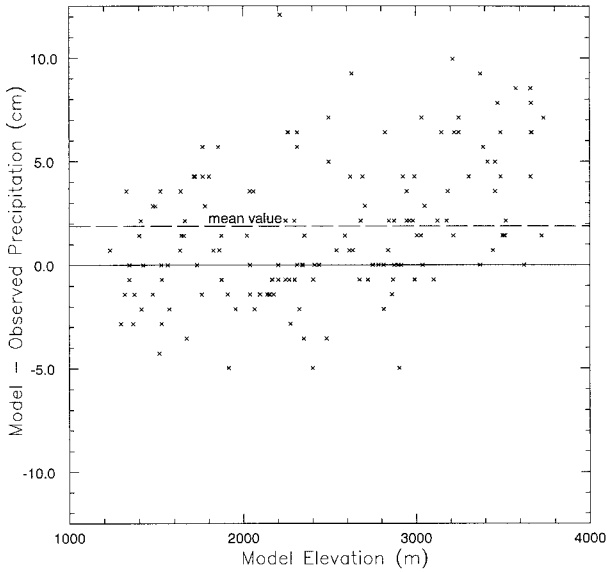

FIG. 12. Total model precipitation at each climatological station, as defined in Fig. 9, minus the total observed precipitation at that station, plotted against the elevation of the nearest model grid point. Dashed line represents the mean value of the precipitation difference for these stations, $y = 1.88$ cm. Standard deviation of the difference is 3.60 cm.

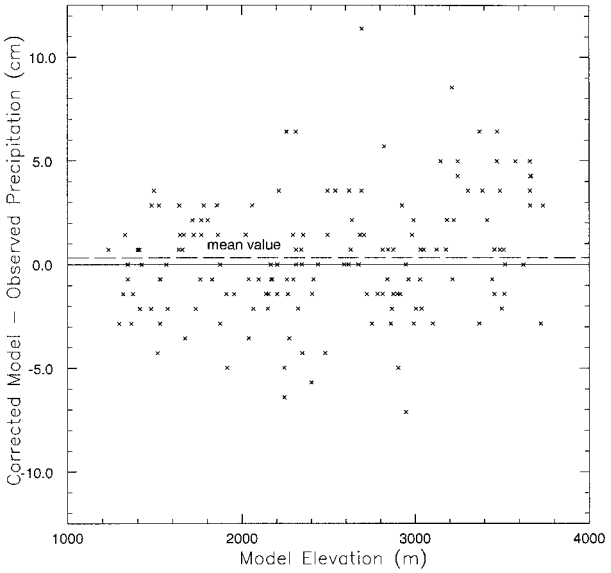

FIG. 13. Same as Fig. 12, but using total model precipitation corrected for the actual station elevation using Eq. (3) and Figs. 9–11. Dashed line represents the mean value of the corrected precipitation difference for these stations, $y = 0.33$ cm. Standard deviation of the difference is 2.87 cm.

three groups and plots the difference between model and observed snowfall. Superimposed on the general scatter is the clear trend toward model overprediction, especially at the higher elevations. Figure 13 uses model output after being corrected for actual station elevation using (3). Though much scatter remains, it can be seen that most of the model bias is removed, and there is little change in model error until about 3400 m, where there still seems to be a slight trend toward model overprediction.

Can this procedure be used for the SNOTEL stations? Figure 14 shows that the numerical model treats most of these sites as if they were in the Colorado–Rio Grande group, whereas six stations seem to belong with the ''outliers'' above the curve in the Platte basin group. These respective basin curves were used to correct the model SNOTEL predictions. As mentioned previously, the model elevations around SNOTEL sites were too low as often as they were too high, and the correction makes very little change in the slight underprediction bias of the model. The SNOTELs that have their precipitation overpredicted are lower than 3000 m and are generally close geographically to the similarly overpredicted climate stations of comparable elevations in the Platte basin, which suggests that the same physiographic feature, or abnormal model forecast, is responsible for both.

c. Gauge errors

There remains a noticeable underprediction of SNOTEL-observed precipitation at sites above 3000 m that is not reflected in the climate station record at this altitude. It is possible that this difference could be due to the underestimation of precipitation by the climate station gauges (see section 5).

A test was done to attempt to quantify the climate station undercatch. One can find a constant b, such that the product of b with the observed monthly precipitation for each station minimizes the least squares error with the model-predicted monthly precipitation over the

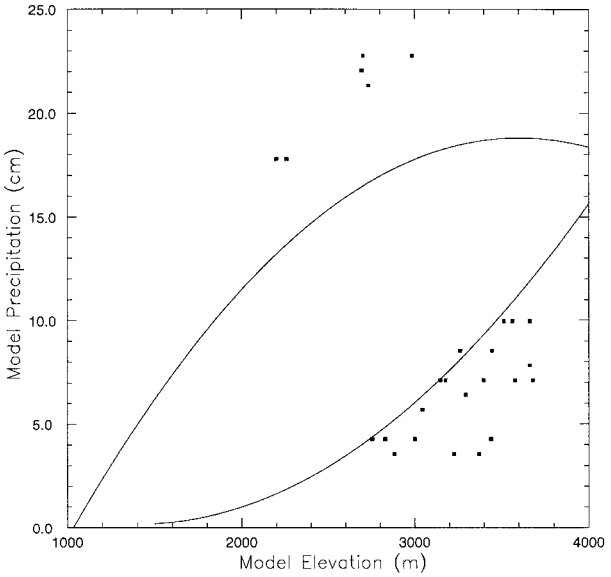

FIG. 14. Same as Fig. 9, but for all SNOTEL stations used in study. Curves are taken from those on Figs. 10 and 11.

TABLE 4. Error statistics and SS between various adjusted model and observed precipitation data. Skill score is that defined in the text: *n*, number of stations in each set; rms, root-mean-square error between model and observed; σ_{obs}, standard deviation of observed precipitation data, and; *b*, adjustment factor for observed precipitation as defined in text. Along first column ordinate, † indicates that observed precipitation has been undercatch adjusted; ‡ indicates that model precipitation has been elevation adjusted. Units of third and fourth columns are mm of precipitation per 28-day period.

Station set	*n*	Rms	σ_{obs}	SS	*b*
Climate	167	4.06	3.25	−0.561	1.260
Climate †	167	3.70	4.10	0.186	
Climate ‡	167	2.89	3.25	0.209	1.002
Climate †‡	167	2.89	3.26	0.214	
High climate	28	5.99	4.31	−0.932	1.407
High climate †	28	4.95	6.07	0.335	
High climate ‡	28	3.72	4.31	0.255	1.135
High climate †‡	28	3.54	4.90	0.478	
SNOTEL	32	6.16	4.66	−0.747	0.751
SNOTEL †	32	5.27	3.50	−1.267	
SNOTEL ‡	32	6.49	4.66	−0.940	0.727
SNOTEL †‡	32	5.47	3.39	−1.604	

whole set of stations. If the model is accurate in the mean, then *b* can be considered the fraction of precipitation actually recorded by a set of observing stations on average.

Table 4 shows the results of this calculation, as well as the skill of various adjusted forecasts. This is expressed in terms of the skill score SS, defined as (Wilks 1995)

$$\mathrm{SS} = 1 - \left(\frac{\mathrm{rms}}{\sigma_{\mathrm{obs}}}\right)^2, \tag{4}$$

where rms is the root-mean-square error, $[\Sigma\,(p_{\mathrm{mdl}} - p_{\mathrm{obs}})^2/n]^{0.5}$, and σ_{obs} is given by $[\Sigma\,(\overline{p} - p_{\mathrm{obs}})^2/n]^{0.5}$, with $\overline{p}$ being the average of the observed monthly precipitation.

For the set of climate stations, *b* turned out to be 1.260, suggesting a 21% undercatch. If all the climate station observations were then multiplied by 1.260, forecasting skill was enhanced (the skill score increased from −0.561 to 0.186). But when just the elevation correction procedure was used, the skill increase was greater (to 0.209), and the value of *b* required was reduced to 1.002.

If only the set of climate stations higher than 2500 m was considered (to coincide with the elevation range of the SNOTELs), the differences were even more dramatic: the skill scores were −0.932 for the uncorrected model, which improved to 0.335 if a 29% undercatch was accounted for. However, even better results (SS = 0.478) were obtained if one first corrected for elevation, then accounted for a 12% undercatch. For the SNOTELs, neither correcting the model output for elevation nor multiplying the observed precipitation by a factor was able to increase the (negative) skill score.

These results suggest that, for the climate stations, elevation overestimation is the major contributor to model overprediction, especially for the lower elevation sites, where any undercatch in the observed seems to be lost in the general scatter of observed precipitation amounts. For the climate stations above 2500 m, the results are most consistent with overprediction being due to the combination of elevation errors and about a 12% undercatch in the observing gauges. Surveys have found that the average undercatch of precipitation at the relatively sheltered (and low altitude) Fort Collins climate station is about 10% (N. Doesken 1998, personal communication). Undercatch amounts of 29% are possible for wind speeds around 3 mi h^{-1} at gauge level (Goodison 1978; Doesken and Judson 1997), but this high of a wind speed at gauge level is believed to be unusual for the western valleys of Colorado (N. Doesken 1998, personal communication).

It is also possible that more precipitation is missed by both the model and the climate stations and is only recorded at the SNOTEL stations. If one assumes that there is about a 15% underprediction in precipitation due to spinup problems, the estimate of *b* moves closer to unity for the SNOTELs and increases by 10%–20% for the climate station set, which is physically plausible. But the skill at the climate stations always improved when elevation correction was performed prior to adjusting the total bias, and the elevation correction reduced the bias adjustment required, unless the stations were missing an implausibly large amount of precipitation. On the other hand, no correction made much of an improvement in the SNOTEL forecasting skill. Probably more resolution in the forecast model would be required to increase forecasting ability; more resolution in the SNOTEL data network would also help to better quantify forecasting skill in these regions.

One factor that is essentially a failure of the model resolution will be addressed here—convective snowfall, which can be a significant contribution to precipitation in regions with steep orography (Reinking and Boatman 1986). Since the finest model grid had 16-km spacing, convection could not be adequately resolved, and the result would be an underestimation in precipitation over the elevated terrain where winter season convection is common. To see if this is a plausible hypothesis, the 0000 UTC lapse rate from 700 mb (roughly 3000 m) to 500 mb at Denver was found for days with significant precipitation and compared to the total SNOTEL precipitation recorded for the 24-h period containing the sounding time. Table 5 shows the temperature difference in degrees Celsius and the corresponding ratio of observed SNOTEL precipitation to model precipitation at the nearest grid points. A difference of about 25°C would correspond to a dry-adiabatic lapse rate; a temperature difference near this value would imply that conditions are favorable for mountain convection. The table shows that when this parameter exceeds about 18°C, precipitation is always underpredicted. When the temperature difference is less than this value, usually there is overprediction.

TABLE 5. Difference between 700-mb temperatures and 500-mb temperatures at Denver from model initialization, in °C. Third column is ratio of observed SNOTEL precipitation for major events to microphysics-predicted precipitation for the daily period including the initialization time shown.

Analysis time (UTC)	$T_{700} - T_{500}$	p_{obs}/p_{micro}
0000 22 Apr	16	0.47
0000 27 Apr	15	0.51
0000 18 Apr	17	0.68
0000 23 Apr	18	0.86
0000 11 Apr	16	0.97
0000 17 Apr	17	0.98
0000 10 Apr	17	1.06
0000 19 Apr	18	1.93
0000 24 Apr	21	2.21
0000 16 Apr	21	2.85
0000 30 Apr	16	3.10
0000 26 Apr	21	3.30
0000 25 Apr	23	4.64
0000 12 Apr	15	7.16
0000 9 Apr	24	460

To take the most extreme example, at 0000 UTC 9 April there was a very steep lapse rate over Colorado in advance of a cold front that would bring heavy precipitation east of the mountains by the morning of 10 April. However, on this particular day (8 April LT) the cooling remained aloft, and in the absence of upslope the precipitation was confined to the mountainous areas, which is a typical Colorado scenario on winter convection days. The model produced virtually no precipitation for the simulation starting 0000 UTC 8 April, or for the first 12 h of the 0000 UTC 9 April simulation. And, finally, the Idarado SNOTEL in the San Juan Mountains recorded no precipitation on the days of 8–10 April, but 1.1 in. (2.8 cm) of water-equivalent precipitation on 9 April; such large rates of snowfall again suggest that convection was responsible.

8. Conclusions

The addition of a bulk microphysics scheme to a mesoscale forecasting model can improve a model's QPF skill. The microphysics model is better at predicting the occurrence or nonoccurrence of precipitation, whereas the use of a ''dump-bucket'' scheme tends to overestimate the area of regions receiving threshold amounts of precipitation. The microphysics model also better reproduces the maximum amounts of solid precipitation that can be expected from a winter-type event; the dump-bucket model tends to underestimate this amount. Thus the use of the bulk microphysics produces improved forecasting over timescales on the order of a month, which would make it useful for snowpack prediction. However, the microphysics forecasting model does less well at specifying the precise geographic location of precipitation features than at predicting the features themselves.

The use of a community-based observational station network causes a systematic bias in comparing model to observed precipitation, because the model uses an average height over a whole grid cell, while the community is more often than not below this height. The result is an overestimation of the total precipitation at such stations. The use of an extrapolation based on model output in the appropriate geographical area helped to remove much of this bias, and it is possible that more geographical decompositions might increase the skill of the procedure even more.

There is evidence that the microphysics forecasting model is causing more precipitation than is observed at the climate stations and lower SNOTELs, and less than is observed at SNOTELs located above 3000 m. This could be due to either problems with the climate stations rain gauges not recording all of the precipitation, or to failures of the forecast model to resolve the terrain, or the atmospheric processes (such as convection).

The improvement in total precipitation forecasting by the use of bulk microphysics is most noticeable when a SNOTEL station set is used to verify the model, though the resolution of the model is too low to show much skill in forecasting for individual SNOTELs or individual snow events. A denser operational SNOTEL network would allow a better assessment of forecasting skill at these locations. The use of improved upper-air and other data analyses could lead to forecasts of higher quality for all meteorological variables. Further improvements in bulk microphysics precipitation predictions using this forecast model would probably require improvements in the cloud radiation scheme, improvement in surface layer schemes and other influences on the behavior of arctic air masses, improvement in convective schemes, and the reduction of model grid spacing to a few kilometers.

Acknowledgments. This work was supported by the Colorado Agricultural Experiment Station under Grant COL0692, the National Science Foundation under Grant ATM-9420045, and the Cooperative Program for Operational Meteorology under Contract UCAR S96-71867. The data used for the model simulations were provided by NCEP and the Alden Corporation. John Kleist and Nolan Doesken of the Colorado Climate Center assisted in obtaining the SNOTEL data and in the interpretation of the data. The help of Brenda Thompson and Abby Hodges in preparing this manuscript is greatly appreciated, as are the comments of two anonymous reviewers.

APPENDIX

Set-Based Skill Scores

The bias score B is given by

$$B = \frac{F}{O}, \tag{A1}$$

where F is the number of stations in which the model predicts precipitation will reach or exceed a certain threshold and O is the number of stations where precipitation was observed to exceed the same threshold. So the bias score is a way to determine if areal coverage is overforecast or underforecast, while the matching of features' location is not considered.

The threat score T is defined to be

$$T = \frac{\mathrm{CF}}{F + O - \mathrm{CF}}, \tag{A2}$$

where CF is the number of correctly forecast stations (both model and observations produce precipitation at or above a given threshold), and F and O are as defined above. Stated in terms of sets, CF is the intersection of sets F and O, and the threat score is the ratio of the intersection of F and O to their union. The largest possible threat score then is 1.0, when F and O are exactly the same set of stations. The least possible threat score is 0.0, which occurs when no stations are correctly forecast and either F or O is nonempty. A good bias score (near 1) but a low threat score implies that areal extent is well represented by the model but geographical location is not. When the bias score or its reciprocal is equal to the threat score, then one of either the forecast precipitation set or observed precipitation set is a subset of the other.

There are a few possible disadvantages associated with the use of the threat score. One is that the correct forecasting of the nonoccurrence of an event (e.g., precipitation not achieving a certain threshold) is not taken into account (Doswell et al. 1990). Another disadvantage is that a certain threat score, depending on the number of forecast and observing stations, can be achieved simply by chance and so represents no true skill. But this threshold threat score cannot be known from just the actual threat score (Mason 1989).

Doswell et al. advocate the use of the HSS as the score of choice that takes into account both expected skill based on chance and the correct forecasting of the nonoccurrence of events. This score is defined as

$$\mathrm{HSS} = \frac{C - E}{N - E}, \tag{A3}$$

where C is the number of correct forecasts both of precipitation and of nonprecipitation [so $C = \mathrm{CF} + (N - F - O + \mathrm{CF})$]. Here, E is the number of correct forecasts expected due purely to chance, given by

$$E = \frac{O \times F + (N - O) \times (N - F)}{N}. \tag{A4}$$

REFERENCES

Anthes, R. A., 1983: Regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.,* **111,** 1306–1335.

——, Y.-H. Kuo, E.-Y. Hsie, S. Low-Nam, and T. W. Bettge, 1989: Estimation of skill and uncertainty in regional numerical models. *Quart. J. Roy. Meteor. Soc.,* **115,** 763–806.

Arakawa, A., and V. R. Lamb, 1981: A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Wea. Rev.,* **109,** 18–36.

Barnes, S. L., 1973: Mesoscale objective map analysis using weighted time series observations. NOAA Tech. Memo. ERL NSSL-62, 60 pp. [NTIS COM-73-10781.]

Beitler, B. A., 1994: Mesoscale numerical prediction of Colorado snowfall and winds. M.S. thesis, Dept. of Atmospheric Science, Colorado State University, Fort Collins, CO, 84 pp. [Available from Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Bleck, R., and S. G. Benjamin, 1993: Regional weather prediction with a model combining terrain-following and isentropic coordinates. Part I: Model description. *Mon. Wea. Rev.,* **121,** 1770–1785.

Clark, T. L., and R. D. Farley, 1984: Severe downslope windstorm calculations in two and three spatial dimensions using anelastic interactive grid nesting: A possible mechanism for gustiness. *J. Atmos. Sci.,* **41,** 329–350.

Colle, B. A., and C. F. Mass, 1996: An observational and modeling study of the interaction of low-level southwesterly flow with the Olympic Mountains during COAST IOP 4. *Mon. Wea. Rev.,* **124,** 2152–2175.

Cotton, W. R., G. Thompson, and P. W. Mielke, 1994: Real-time mesoscale prediction on workstations. *Bull. Amer. Meteor. Soc.,* **75,** 349–362.

——, J. F. Weaver, and B. A. Beitler, 1995: An unusual summertime downslope wind event in Fort Collins, Colorado, on 3 July 1993. *Wea. Forecasting,* **10,** 786–797.

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

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson, 1986: Biosphere–Atmosphere Transfer Scheme for the NCAR Community Climate Model. Tech. Rep. NCAR/TN-275+STR, NCAR, Boulder, CO, 69 pp.

Doesken, N., and G. Schaefer, 1987: The contribution of SNOTEL precipitation measurements to climate analysis, monitoring, and research in Colorado. *Proc. Western Snow Conf., 55th Annual Meeting,* Vancouver, BC, Canada, Western Snow Conference, 20–30.

——, and A. Judson, 1997: *The Snow Booklet: A Guide to the Science, Climatology, and Measurement of Snow in the United States.* 2d ed. Colorado Climate Center, Department of Atmopheric Science, Colorado State University, 86 pp.

Doswell, C. A., III, R. Davies-Jones, and D. K. Leller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting,* **5,** 576–585.

Flatau, P. J., G. J. Tripoli, J. Verlinde, and W. R. Cotton, 1989: The CSU-RAMS cloud microphysical module: General theory and code documentation. Atmospheric Science Paper 451, 88 pp. [Available from Dept. of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Gal-Chen, T., and R. C. J. Somerville, 1975: On the use of a coordinate transformation for the solution of the Navier–Stokes equations. *J. Comput. Phys.,* **17,** 209–228.

Gartner, W. E., M. E. Baldwin, and N. W. Junker, 1996: Regional analysis of quantitative precipitation forecasts from NCEP's early Eta and meso-Eta Models. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 169–171.

Giorgi, F., and G. T. Bates, 1989: The climatological skill of a regional model over complex terrain. *Mon. Wea. Rev.,* **117,** 2325–2347.

Goodison, B. E., 1978: Accuracy of Canadian snow gauge measurements. *J. Appl. Meteor.,* **27,** 1542–1548.

Hoke, J. E., N. A. Phillips, G. J. DiMego, J. T. Tuccillo, and J. G. Sela, 1989: The Regional Analysis and Forecast System of the National Meteorological Center. *Wea. Forecasting,* **4,** 323–334.

Jarraud, M., A. J. Simmons, and M. Kanamitsu, 1987: The concept,

implementation and impact of an envelope orography. *Proc. Seminar/Workshop 1986 on Observation, Theory and Modeling of Orographic Effects*, Vol. 2, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 81–127.

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

——, ——, B. E. Sullivan, K. F. Brill, and F. J. Hughes, 1992: Seasonal and geographic variations in quantitative precipitation prediction by NMC's Nested-Grid Model and Medium-Range Forecast Model. *Wea. Forecasting,* **7,** 410–429.

Klemp, J. B., and R. B. Wilhelmson, 1978: Simulations of right- and left-moving storms produced through storm-splitting. *J. Atmos. Sci.,* **35,** 1097–1110.

Manikin, G. S., K. E. Mitchell, and E. Rogers, 1996: Severe weather forecasts using the Eta and meso Eta models. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 295–296.

Mason, I., 1989: Dependence of the critical success index on sample climate and threshold probability. *Aust. Meteor. Mag.,* **37,** 75–81.

McMillan, G. D., 1981: SNOTEL: A management tool for the future. *Proc. Western Snow Conference, 49th Annual Meeting,* St. George, UT, Western Snow Conference, 116–119.

Mesinger, F., Z. I. Janjić, S. Ničković, D. Gavrilov, and D. 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.

Meyers, M. P., P. J. DeMott, and W. R. Cotton, 1992: New primary ice nucleation parameterizations in an explicit cloud model. *J. Appl. Meteor.,* **31,** 708–721.

NCDC, 1995: *Climatological Data.* Vol. 100, No. 4. [Available from National Climatic Data Center, Asheville, NC 28801.]

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

Rangno, A. L., and P. V. Hobbs, 1994: Ice particle concentrations and precipitation development in small continental cumuliform clouds. *Quart. J. Roy. Meteor. Soc.,* **120,** 573–601.

Reinking, R. F., and J. F. Boatman, 1986: Upslope precipitation events. *Mesoscale Meteorology and Forecasting,* P. S. Ray, Ed., Amer. Meteor. Soc., 437–471.

Rhea, J. O., 1978: Orographic precipitation model for hydrometeorological use. Ph.D. dissertation, Colorado State University, Fort Collins, CO, 199 pp. [Available from Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, 1996: Changes to the operational ''early'' eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting,* **11,** 391–413.

Schultz, P., 1995: An explicit cloud physics parameterization for operational numerical weather prediction. *Mon. Wea. Rev.,* **123,** 3331–3343.

Snook, J. S., and R. A. Pielke, 1995: Diagnosing a Colorado heavy snow event with a nonhydrostatic mesoscale numerical model structured for operational use. *Wea. Forecasting,* **10,** 261–285.

Spreen, W. C., 1947: A determination of the effect of topography upon precipitation. *Trans. Amer. Geophys. Union,* **28,** 285–290.

Thompson, G., 1993: Prototype real-time mesoscale prediction during 1991–92 winter season and statistical verification of model data. M.S. thesis, Department of Atmospheric Science, Colorado State University, Fort Collins, CO, 105 pp. [Available from Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Tremback, C. J., 1990: Numerical simulation of a mesoscale convective complex: Model development and numerical results. Ph.D. dissertation, Atmospheric Science Paper 465, Dept. of Atmospheric Science, Colorado State University, Fort Collins, CO, 247 pp. [Available from Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523.]

Tripoli, G. J., and W. R. Cotton, 1981: The use of ice-liquid water potential temperature as a thermodynamic variable in deep atmospheric models. *Mon. Wea. Rev.,* **109,** 1094–1102.

——, and ——, 1982: The Colorado State University three-dimensional cloud/mesoscale model—1982. Part I: General theoretical framework and sensitivity experiments. *J. Rech. Atmos.,* **16,** 185–220.

Walko, R. O., W. R. Cotton, J. L. Harrington, and M. P. Meyers, 1995: New RAMS cloud microphysics parameterization. Part I: The single moment scheme. *Atmos. Res.,* **38,** 29–62.

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