FORECASTER'S FORUM

Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations

PAUL J. ROEBBER

Atmospheric Science Group, Department of Mathematical Sciences, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

DAVID M. SCHULTZ

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, Oklahoma

BRIAN A. COLLE

Institute for Terrestrial and Planetary Atmospheres, State University of New York at Stony Brook, Stony Brook, New York

DAVID J. STENSRUD

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

16 October 2003 and 3 May 2004

ABSTRACT

A large gap in skill between forecasts of the atmospheric circulation (relatively high skill) and quantitative precipitation (low skill) has emerged over the past three decades. One common approach toward closing this gap has been to try to simulate precipitation features directly by decreasing the horizontal grid spacing of the numerical weather prediction models. Also at this time, research has begun to explore the benefits of short-range ensemble forecast methods. The authors argue that each approach has benefits: high-resolution models assist in the development of a forecaster's conceptual model of various mesoscale phenomena, whereas ensembles help quantify forecast uncertainty. A thoughtful implementation of both approaches, in which this complementary nature is recognized, will improve the forecast process, empower human forecasters, and consequently add value relative to current trends. The science and policy issues that must be addressed in order to maximize this forecast potential are discussed.

1. The forecast skill gap

Because of the relative lack of skill in forecasts of precipitation [see the National Center for Atmospheric Research (NCAR) Web site for recent verification statistics for its models online at http://sgi62.wwb.noaa.gov:808/STATS/STATS.html (height and wind) and at http:www.hpc.ncep.noaa.gov/html/hpcverif.html (precipitation)], both the U.S. Weather Research Program (Fritsch et al. 1998) and the National Research Council Board on Atmospheric Sciences and Climate (National Research Council 1998, p. 174) have declared a principal research goal to be improved forecasts of precipitation occurrence and amount. Numerical weather prediction (NWP) is the main vehicle for forecasting at mesoscale spatial and temporal scales. Improving NWP forecasts can involve increasing the amount of data and advancing data assimilation, improving model physical parameterizations, resolution, and model postprocessing techniques, and addressing inherent forecast uncertainties using ensembles or other appropriate probabilistic methods. The purpose of this paper is to discuss how to improve operational NWP through high-resolution (section 2) and ensemble (section 3) forecasting. Nevertheless, advances in data, data assimilation, model formulation, parameterization, and model postprocessing are intimately connected and are also discussed, as are the limitations imposed by operational realities (section 4). Discussion of the complementary nature of high-resolution and ensemble approaches is provided in section 5, and the paper concludes with our recommendations (section 6).

Corresponding author address: Paul J. Roebber, Department of Mathematical Sciences, University of Wisconsin—Milwaukee, 3200 N. Cramer Ave., Milwaukee, WI 53211.
E-mail: roebber@uwm.edu

© 2004 American Meteorological Society

2. High-resolution models

The National Centers for Environmental Prediction (NCEP) has decreased model grid spacing in their models from 190.5-km grid spacing during the 1970s [Limited-area Fine-mesh Model (LFM)]; Petersen and Stackpole 1989) to 12 km presently in the Eta model (Black 1994) to 8 km for the Nonhydrostatic Mesoscale Model (Janjic et al. 2001). Increasing resolution has resulted in improved model simulations and predictions of key atmospheric phenomena, such as rapidly developing extratropical cyclones in the western Atlantic (e.g., Kuo and Low-Nam 1990; Uccellini et al. 1999), Rocky Mountain lee cyclogenesis (Schultz and Doswell 2000, p. 153), cold surges east of the Rockies (Mesinger 1996), Appalachian cold-air damming (e.g., Weygandt and Seaman 1994), freezing precipitation (e.g., Roebber and Gyakum 2003), orographic winds and precipitation (e.g., Mass et al. 2002 and references within), sea/lake-breeze circulations (e.g., Manobianco and Nutter 1999; Roebber and Gehring 2000), lake-effect snowstorms (e.g., Ballentine et al. 1998; Steenburgh and Onton 2001), and convective systems (e.g., Weisman et al. 1997; Bernadet et al. 2000; Nielsen-Gammon and Strack 2000; Roebber et al. 2002).

a. Data limitations

While synoptic-scale datasets serve regional-scale (i.e., grid spacing that is considerably coarser than high resolution, for the purposes of this paper, 20 km or more and 10 km or less, respectively) forecast models sufficiently well in many situations, the frequent and sometimes spectacular failures associated with the limitations of existing observing networks (e.g., Langland et al. 2002; Zupanski et al. 2002; Bosart 2003; McMurdie and Mass 2004) have led to repeated calls for increased and better use of observations (e.g., Bosart 1990; Emanuel et al. 1995; Douglas and Stensrud 1996; Schlatter and Lord 1997). High-resolution models may be even more sensitive than regional models to data limitations. For example, Zhang et al. (2002) showed that withholding even a single sounding can significantly alter the detailed mesoscale distribution of forecast precipitation in a major cyclogenesis event. Steenburgh and Onton (2001) found that a Great Salt Lake–effect snow event was sensitive to lake temperatures and upstream humidity. Roebber et al. (2002) showed that a major limitation to skillful 24-h forecasts of the 3 May 1999 tornado outbreak was the lack of upstream synoptic-scale data.

An explicitly stated goal of the Center for Analysis and Prediction of Storms at the University of Oklahoma is to predict specific convective storms (e.g., Lilly 1990). The limited progress toward this goal thus far (Weygandt et al. 2002) suggests that increased mesoscale data may be required, such as might be obtained in part from the nationwide distribution of the Level-II Weather Surveillance Radar-1988 Doppler (WSR-88D) radar data (Droegemeier et al. 2002). Even with these increased data assets, success in that endeavor is far from certain (e.g., Gallus and Segal 2001). Such efforts serve as a warming of the observational demands for future improvement of high-resolution NWP.

b. Does higher resolution lead to greater skill?

Using 2 yr of real-time output from the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5; Dudhia 1993; Grell et al. 1994), Mass et al. (2002) showed that there was a significant improvement in skill of wind direction, temperature, and precipitation as model grid spacing was reduced from 36 to 12 km. Therefore, some resolution over complex topography is needed to properly simulate the basic flow structures and precipitation enhancements. There is a point of diminishing returns, however, since a decrease in grid spacing over western Washington from 12 to 4 km showed little additional improvement for wind, temperature, and precipitation amounts less than 2 in. $(24\ \mathrm{h})^{-1}$ (Colle et al. 2000). In areas of lower topography such as the northeastern United States, however, there was little improvement in quantitative precipitation forecasts when grid spacing was decreased from 36 to 12 km (Colle et al. 2003). Similarly, Gallus (2002) found a degradation of skill for midwestern convective systems when grid spacing was reduced from 30 to 10 km.

c. Improved physics

A general trend over the years is that as model grid spacing becomes smaller, the need for more sophisticated physical process parameterization schemes increases. For example, if high-resolution models are to correctly forecast the development and evolution of deep convection under quiescent large-scale forcing, then what physical processes need to be accurately reproduced within the model? The planetary boundary layer often sets the stage for the development of convection, so an accurate depiction of this layer would be needed. An accurate depiction of the planetary boundary layer, however, requires accuracy in the representation of all of the following physics: incoming shortwave and outgoing longwave radiation, ground heat flux, sensible and latent heat flux, turbulent mixing, and latent heating due to condensation. The radiation calculations require knowledge of the cloud field in terms of liquid water path, percent cloud coverage at numerous atmospheric levels, aerosols, cloud-base and cloud-top temperatures, and ground surface temperature. The ground, sensible, and latent heat fluxes each require knowledge of the soil type, soil moisture, soil temperature with depth, and vegetation type, vegetation health, and vegetation coverage. The calculation of turbulent mixing requires knowledge of the stability, vertical wind shear, eddy sizes, vertical motions, boundary layer circulations, and entrainment processes.

Marshall et al. (2003) document many of the challenges to accurate forecasts of the planetary boundary layer, and indicate that all of the present parameterization schemes that help determine boundary layer evolution have inaccuracies. Other important physics limitations exist as well, including representation of cloud microphysics (Colle and Mass 2000; Larson et al. 2001; Lynn et al. 2004a,b). Physical processes appear to get more complex and interrelated as we move toward smaller grid spacing, and we lack the needed observations to define many of the required variables. Can we reasonably simulate all this complexity with sufficient accuracy to yield increasingly skillful forecasts? This is an important question, but we still fall short of having a good answer.

d. Remaining questions

As suggested by the above discussion, there are a number of important questions that remain to be resolved regarding the use of high-resolution models in operations.

- *What is the optimal grid spacing and nesting strategy for high-resolution models?* Numerous studies show that forecasts are not always the best at the smallest grid spacing. How do we decide when the point of diminishing returns has been reached? Nested grids are still needed for high-resolution modeling, but recent studies have shown that parameterized convection in the outer domain can have a detrimental impact on the explicit precipitation forecasts within the inner nest (Warner and Hsu 2000). What nested grid size is needed to mitigate this problem?
- *How do we verify high-resolution model forecasts and compare them against lower-resolution forecasts?* If we want to make rational decisions regarding the value of high-resolution forecasts, then we must find ways to verify these forecasts and compare them to forecasts at different resolutions in ways that measure the true skill and value of these forecasts.
- *How do we initialize variables that we do not observe?* Many of the variables needed in sophisticated physical process schemes are not commonly observed (e.g., soil moisture). Similar problems exist with soil temperatures, water temperatures, cloud variables, vegetation parameters, aerosols, and many others.
- *How do we disseminate high-resolution forecast data?* Currently, the 12-km Eta Model forecasts are provided to forecasters on a nominal 40-km grid. If forecasts are produced at 2 km, then how do we get the full grid information to forecasters? What tools exist for them to examine these data quickly? At what time frequency will output be needed?
- *What model parameterizations need further development in order to be properly applied at high resolution?* Many large mesoscale forecast errors still develop from fundamental deficiencies in boundary layer, convective, and microphysical parameterizations.
- *How can high-resolution forecasts be applied in the context of probabilistic forecasting?* High-resolution forecasts have the advantage of being able to resolve important mesoscale flows and precipitation, yet a challenge is to put them in the context of the predictability on any given day.

3. Ensembles

In addition to using high-resolution forecast models for short-range weather forecasting, another approach that has shown great promise is ensembles of lower-resolution models. A few studies have found that ensembles of lower-resolution models provide greater skill than single forecasts of higher-resolution models, when verified over a large sample of events (e.g., Stensrud et al. 1999; Wandishin et al. 2001; Grimit and Mass 2002). For example, for a 6-month period over the Pacific Northwest, Grimit and Mass (2002) demonstrated that a five-member MM5 ensemble at 12-km grid spacing was as skillful as a single deterministic 4-km run. Mullen and Buizza (2002) have also shown that coarser-resolution (total wavenumbers 159 and 255) larger-member ensembles can outperform higher-resolution (though still coarse-grain with total wavenumber 319), smaller-member ensembles in specific situations.

A primary advantage of ensembles is that they are inherently probabilistic for all forecast fields and so can express uncertainty directly (e.g., Tracton and Kalnay 1993; Palmer 2002). Hence, users can make informed decisions based on these probabilities and their own cost/loss ratios. For example, Palmer (2002) discussed the extratropical cyclone that devastated parts of Europe on 26 December 1999. Although the single deterministic prediction from the European Centre for Medium-Range Weather Forecasts (ECMWF) did not develop that storm, about a third of the ECMWF Ensemble Prediction System's 50 members produced intense cyclogenesis, with a probability of gusts exceeding 40 m s^{-1} over 30% in some locations (Fig. 3 and 4 in Palmer 2002).

Ensemble techniques have been developed that vary the initial conditions, physical parameterizations, or numerical models, or are combinations of two or more of the above methods. These techniques all seek to increase forecast skill by combining independent information obtained from individual ensemble members. However, a number of important questions remain to be resolved.

- *What is the best way to construct an ensemble?* For example, there are several ways to vary initial conditions and the optimal method for accomplishing this has not been definitively determined.
- *What is the relative role of initial conditions versus*

model formulation in constructing ensembles? While the impact of data limitations is implicit in the ensemble modeling approach through probabilistic forecasting, this approach is not a panacea. The net utility of such forecasts is drastically reduced where the initial data are insufficient to provide any sharpness in the probability distributions. Ensembles typically use lower-resolution models with more simplified physics packages (e.g., parameterized versus explicit convection), thus the role of model errors in constructing ensembles is likely to be important, as well.
- *For what temporal scales are ensembles best suited?* Most typically, ensembles have been used for medium-range and longer time scales (e.g., Palmer 2002), although Stensrud et al. (1999) and Elmore et al. (2002) have presented evidence that short-range forecasting can also benefit from ensemble approaches.
- *What is the best way to produce probabilistic forecasts from the ensemble output?* Although ensembles are well designed for generating probabilistic forecasts, the accuracy of these forecasts can be compromised by model biases; thus, model output may need to be bias corrected before producing the probabilistic forecasts. Weighting of ensemble members using various statistical processing techniques likely would also prove useful.
- *What is the source of the underdispersion of ensemble system, and how can this best be corrected?* The envelope of solutions generated by an ensemble system is not, in all cases, sufficiently large to encompass reality. For example, a substantial convective rainfall event may not be captured by any ensemble member, even when mixed model, mixed physics, and variously perturbed initial conditions are used. The inability of the ensemble to forecast the forecast skill of the ensemble mean is believed to be tied to underdispersion (Hamill and Colucci 1998; Stensrud et al. 1999), but increasing dispersion can also lead to a deterioration of predictive skill (Wandishin et al. 2001).

4. Operational realities

In addition to the fundamental science questions outlined in sections 2 and 3, there are a number of procedural issues that must be considered before promising research results can be effectively transferred to operations.

a. Is the model capable?

In the case of high resolution, while it is reassuring to know that models can replicate detailed meteorological structures, it is important to recognize that the use of models to study physical processes and the use of models to make weather forecasts are distinctly different applications of the same tool. A *capable* model is defined as one that can replicate an observed atmospheric feature of interest (e.g., orographic precipitation), and is critical to the study of physical processes. A capable model, however, is not necessarily more useful in the scientific forecast process (Doswell 1986; Doswell and Maddox 1986; Hoffman 1991; Andra et al. 2002; Roebber et al. 2002; see the appendix) than a model with coarser grid spacing that cannot explicitly resolve the same feature. For example, a high-resolution model may be capable, yet lack reliability (e.g., overforecasting the feature of interest), and hence interfere with hypothesis testing. A rigorous comparison of daily model forecasts with observations must be undertaken to determine the model reliability for predicting phenomena of interest. Such studies have been performed for trough passages over the eastern Pacific (Colle et al. 2001), precipitation and surface winds in the Pacific Northwest (Colle et al. 1999, 2000; Mass et al. 2002), lake-breeze circulations on the western shores of Lake Michigan (Roebber and Gehring 2000), and convective occurrence and mode in the upper Midwest (Fowle and Roebber 2003). Even a model that is not capable may still provide useful conceptual information, either directly or through various postprocessing techniques.

b. Postprocessing of model data

When comparing the benefits of differing forecast approaches, it is important to examine both forecast skill and forecast value. One frequently overlooked area of NWP is postprocessing, in which systematic model errors are reduced via a variety of statistical techniques. Statistical postprocessing can increase both forecast skill and forecast value dramatically (e.g., Vislocky and Fritsch 1995; Hall et al. 1999; Koizumi 1999; Peyraud 2001; Krishnamurti et al. 2001; Hart et al. 2004), although Atger (2003) raises concerns about such techniques for ensembles related to sampling limitations. Postprocessing is particularly of interest since its costs in computer time typically are much less than the costs of generating the numerical model forecast. Comparisons should be made between the raw output from high-resolution models and the postprocessed output from lower-resolution models. For example, regression techniques such as Model Output Statistics (MOS; Glahn and Lowry 1972) can account for the effects of complex terrain, such as cold pools in valleys and local thermally driven flows, even if the numerical model upon which the regression is based does not resolve the topographic features (e.g., Hart et al. 2004). Postprocessing approaches need further exploration.

c. How is model utility measured?

An important issue is the means by which forecasters can most profitably use the NWP output from different modeling approaches. This relates directly to a forecast principle, well understood by experienced forecasters, that has remained true since the issuance of the first model guidance: *NWP output cannot be taken literally.* Subelements of a forecast (Pliske et al. 2004) may be predicted well, whereas others will not. How do forecasters know what to believe and what not to believe in the model output? A related consideration is the definition of an appropriate measure of mesoscale predictability. Which is more useful: gridpoint verifications or some measure of conceptual guidance? Are there situations where one measure is more appropriate than others? How is model utility measured?

Gallus (1999) compared forecasts from the operational Eta to a quasioperational version employing the Kain–Fritsch convective parameterization (EtaKF). The EtaKF often produced precipitation with reasonable magnitudes, but offset from the observed locations, while the Eta produced lesser amounts across wider areas. Which is the better forecast? Baldwin et al. (2001) illustrated this verification problem by considering an idealized observed precipitation distribution composed of a large-scale band with smaller embedded maxima (Fig. 1a). A low-resolution forecast where the large-scale precipitation band had the wrong orientation and no smaller maxima (Fig. 1b) was compared to a high-resolution forecast where the large-scale band was well positioned, but embedded maxima were located incorrectly (Fig. 1c). Baldwin et al. (2001) showed that the mislocation of the high-intensity embedded structures penalized the high-resolution forecast, leading to a lower skill score using traditional measures of verification. Consequently, Baldwin et al. (2001) advocate an ‘‘events-oriented’’ approach to verification.

An application of this concept can be found in Fowle and Roebber (2003), who investigated short-range operational forecasts from the University of Wisconsin—Milwaukee 6-km MM5 for April–September 1999. Although the threat scores for measurable precipitation from these forecasts were comparable to that of lower-resolution operational models, Fowle and Roebber (2003) found that the 6-km MM5 output could be used to skillfully assess convective occurrence, timing, and mode (categorized as linear, multicell, or isolated; the true skill statistic for this aspect ranged from 0.86 to 0.91, where 0 and 1 represent no skill and perfect skill, respectively). Similar results concerning high-resolution forecasts of convective mode have been obtained in the Great Plains region in operational runs of the Weather Research and Forecast (WRF) model in support of the Bow Echo and Mesoscale Convective Vortices (MCVs) Experiment (BAMEX; C. Davis 2003, personal communication). These successes indicate that high-resolution simulations are capable of producing new forms of guidance reliably in an operational setting.

This serves to illustrate that as model grid spacing decreases, the need to use the model output to assist in the development of *conceptual understanding* rather than gridpoint (literal) forecasting also increases. This is so because the decreased grid spacing leads to an increase in degrees of freedom and, owing to the present

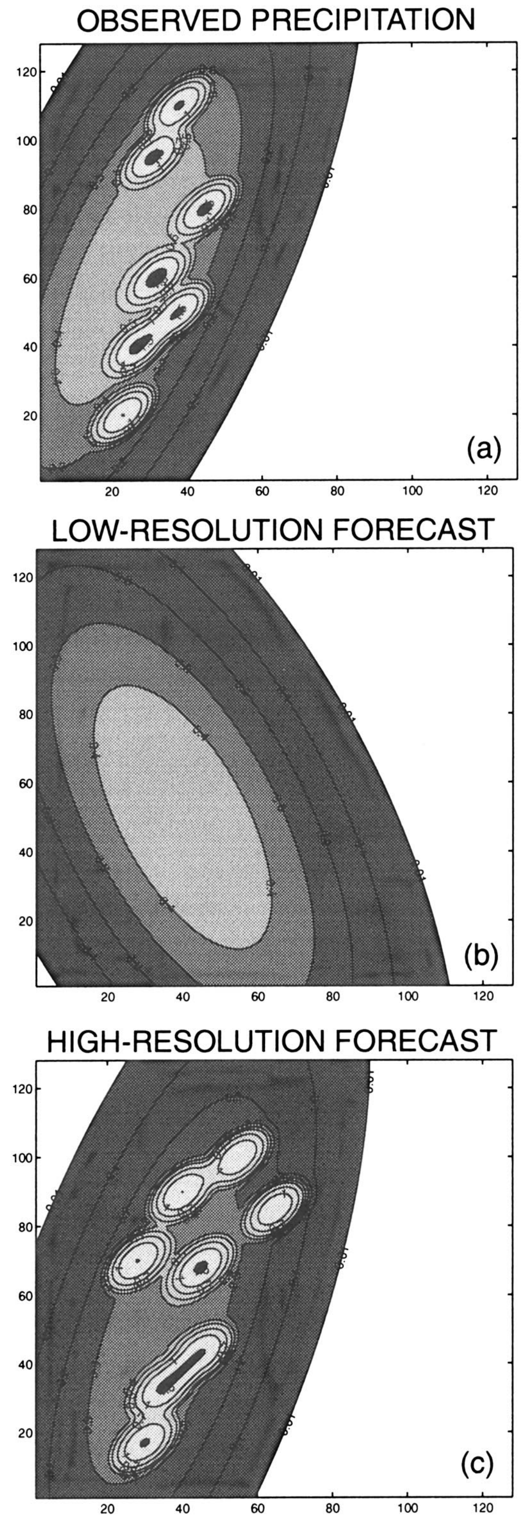

FIG. 1. Example of an idealized precipitation event: (a) observed, (b) low-resolution forecast, and (c) high-resolution forecast. [Adapted from Baldwin et al. (2001).]

large errors in NWP models (e.g., Orrell et al. 2001), an increase in model error as determined by standard gridpoint verification measures. High-resolution models may produce wonderfully detailed, but inaccurate, features, making such forecasts unfit for direct public dissemination. The information from high-resolution models, however, can be a powerful means for refining forecast hypotheses (e.g., Roebber et al. 2002; Schumacher 2003) and can contribute substantially to the scientific forecast process.

Since all model forecasts are imperfect, forecasters need to be able to ascertain when and why forecasts are going wrong. Proper tools (both conceptual understanding and workstation software) for the critical evaluation of forecasts are required and these tools will vary according to the application. For categorical/event forecasts, a basic form of verification is to create a scatterplot of forecast skill in accuracy for events, accuracy for nonevents phase space (McBride and Ebert 2000). Model performance data of this type for specific events (e.g. convection, heavy precipitation, etc.) would assist forecasters in their attempts to construct conceptual models. For example, the model guidance might indicate the likelihood of a significant precipitation event, but this may be tempered by the knowledge that the model skill is derived primarily from the detection of nonevents. In this case, less reliance would be placed upon this guidance. In addition, trends in relationships between large-scale model fields (such as vertical motion forcing and thermodynamics) and small-scale details such as convection may be known and used to correct forecasts (e.g., Jankov and Gallus 2004).

Schumacher (2003) argued that forecasters of the future will require both high- and low-resolution gridded model data. For example, although realistic-looking features may be generated in high-resolution models, their small grid spacing may result in fields of derivative quantities (e.g., advection, frontogenesis, convergence) that are too noisy to effectively aid the forecaster in hypothesis testing (regardless of whether those noisy features are real or not). Overall, the scientific forecasting process, using either high-resolution models or lower-resolution ensembles, is potentially compromised if forecasters are not given the proper tools.

d. Time constraints

Operational time constraints can further reduce the potential usefulness of high-resolution models, since, as the amount of available information increases, potential forecast skill increases at a faster rate than the skill actually achieved (Wright 1974; Heideman et al. 1993), and actual skill may even decline (Stewart et al. 1992). Because a high-resolution model produces more data, it is increasingly difficult for a forecaster to fully evaluate the output in a finite time. Hence, where high-resolution model data are available, it is critical that resources be devoted to improving the *use* of the information rather than simply increasing the supply. The output from such models must be tailored to the needs of the forecasters and tools must be developed that allow for timely exploration of the model data. Forecasters would benefit from knowing how well models are performing recently in order to quantify recent model biases and trends. In addition, there are no easy-access tools that forecasters can use to determine what data was thrown out of the data assimilation system. Such ignorance can lead to forecast failures, such as the ''surprise'' 25 January 2000 snowstorm (Langland et al. 2002; Zupanski et al. 2002; Zhang et al. 2002). Another critical operational constraint relates to logistics: a perfect model is of little use if the output arrives after the forecast is due, a relevant issue considering the resource intensiveness of high-resolution models.

5. Combining high resolution with ensembles

As is clear from the previous sections, there are advantages, disadvantages, and challenges facing high-resolution and ensemble modeling systems (Table 1). By combining these approaches, however, it is possible to overcome some individual disadvantages and benefit from their complementary strengths. One possible method of achieving this, as yet untested, would be to use ensemble forecasts to provide guidance on the most likely scenario, the average scenario, or the most damaging scenario. One could then use that information to provide initial and lateral boundary conditions for a single, high-resolution forecast (e.g., Stensrud et al. 2000; Molteni et al. 2001; Marsigli et al. 2001).

In the future, the means to rerun a model forecast after adjusting the initial conditions (e.g., strengthening a feature in the initial conditions to compensate for a potential lack of data about its intensity; Roebber et al. 2002) or physics (e.g., cloud radiative effects) would be another option that forecasters might have available (e.g., Bright et al. 2003). This type of intervention represents another potentially effective combination of the high-resolution and ensemble forecast strategies.

Even without such targeted use of high-resolution and ensemble methods, the complementary nature of these approaches can provide useful information. Consider the case of Tropical Storm Floyd, which made landfall along the southern North Carolina coast on 16 September 1999. Even though the winds with this system weakened rapidly as it moved up the eastern seaboard, the storm produced 25–40 cm of rain over a 12–24-h period just inland of the coast.

Unfortunately, none of the NCEP operational models were able to simulate the heavy precipitation during Floyd (e.g., the 32-km Eta model produced less than 50% of the observed rainfall over the flooded areas of southern New England). However, recent studies have suggested that the interaction of a tropical cyclone with a preexisting baroclinic zone associated with an approaching midlevel trough can induce a broad region of

TABLE 1. Advantages and disadvantages of high-resolution vs ensemble modeling systems. A list of challenges that must be overcome is also included.

High-resolution modeling:

Advantages:
- Better resolution of all weather phenomena
- Better predictability of some phenomena

Disadvantages:
- Costly in terms of computer resources
- Some model errors may increase at increased resolution

Challenges:
- Improved model parameterizations designed specifically for higher resolutions
- Optimal resolution
- Inadequate traditional model verification measures

Ensembles:

Advantages:
- On average, more skillful forecasts than any individual ensemble member
- Output from an unbiased ensemble interpreted directly in terms of probability forecasts

Disadvantages:
- Diagnosing all of the individual ensemble members not feasible

Challenges:
- Best methods to construct ensembles (e.g., varying initial conditions, varying models)
- Balancing the optimal number of ensemble members vs resolution of individual members
- Underdispersion

Both:

Advantages:
- Complementary approaches provide forecaster with greater opportunity for hypothesis formation and testing

Disadvantages:
- Increased amount of model output

Challenges:
- Forecasters' ability to distinguish what to believe in model output from what not to believe
- Visualization, diagnosis and dissemination of large volumes of model output
- Unknown model climatologies for different weather phenomena
- Forecaster understanding of model climatologies limited by constant updating of models
- Limited quantity of observational data
- Best procedures to construct initial conditions
- Opportunities for postprocessing model data
- How to produce probabilistic forecasts
- Forecaster education
- Transferring research results into operations

slantwise ascent and precipitation several hundred kilometers to the northwest of the cyclone (Harr and Elsberry 2000; Atallah and Bosart 2003). Therefore, if a forecaster can relate the relevant dynamics to the output from high-resolution numerical guidance, one can obtain confidence in his or her conceptual model and anticipate the potential for heavy precipitation. Colle (2003) showed using high-resolution MM5 simulations that a combination of moist symmetric instability below 800 hPa, slantwise neutrality aloft, and strong frontogenesis (enhanced by a horizontal gradient in midlevel latent heating across the coast) produced the narrow, intense precipitation band during the Floyd event.

Colle (2003) showed the importance of increased horizontal resolution on the 36-h MM5 precipitation forecasts of Floyd. The relatively coarse resolution 36-km MM5 generated 15–25 cm of precipitation across northeastern New Jersey to western Connecticut, increasing to 25–40 cm in the 4- and 1.33-km simulations (see his Fig. 4). Although there were biases in the higher-resolution MM5 forecasts, with a general underprediction in the heaviest precipitation area and overpredictions toward the coast, these forecasts provided a reasonable representation of the orientation and magnitude of the flooding event.

Meanwhile, an ensemble of numerical guidance even at coarse resolution can be useful to obtain more or less confidence for the heavy precipitation over a particular region. For example, an ensemble of nine different MM5 runs at 32-km grid spacing (initialized at 0000 UTC 16 September 1999) were completed for Floyd (Fig. 2) using identical initial and boundary conditions from the Eta, but different physics options for the convective parameterization and the PBL. Whereas all of the forecasts underestimated the observed storm total precipitation amounts, most solutions suggested the potential for heavy rainfall near the coast. Thus, the ensemble forecast gives further confidence in the potential heavy precipitation, while the high-resolution MM5 run suggests that the ensemble amounts should be increased by 30%–40% based on its better ability to resolve the precipitation band. The ensemble also illustrates some uncertainty in the location of the rainfall maximum; therefore, forecasters need to assign probabilities to these locations based on ensemble guidance.

The initial condition sensitivity for Floyd was also not negligible. Comparisons of these forecasts to one using the aviation run of the global spectral model (AVN) initial conditions and identical physics (not shown) reveal substantial variations in quantitative precipitation forecasts (QPFs) suggesting that an ensemble representing a mix of models and initial conditions might provide the best representation of the forecast sensitivity for this case.

6. Recommendations

What is the future role of the human forecaster as technologies progress? This issue has been debated at several recent conferences (Cyclone Workshop, Monterey, California, in September 2000; the Ninth Mesoscale Conference, Fort Lauderdale, Florida, in August 2001; the 84th Annual Meeting of the American Meteorological Society, Seattle, Washington, in January 2004), and in the literature (e.g., Schwartz 1980; Doswell et al. 1981; Doswell 1986; Tennekes 1988, 1992; Brooks et al. 1992; Brooks and Doswell 1993; http://webserv.chatsystems.com/~doswell/forecasting/dosfuture.html; McIntyre 1999; Bosart 2003; Mass 2003a,b; Glahn 2003).

Mass (2003a) states: "Humans cannot integrate the primitive equations in their heads and would be hard-pressed to improve upon full-resolution model forecasts that include bias removal. Human improvement upon calibrated probabilistic forecasts based on a mesoscale ensemble system is even more unlikely." Instead, Mass (2003a) argues that forecasters should focus their efforts in the 0–12-h range, citing current human superiority in image interpretation that can be advantageously coupled with physical understanding. There is, however, some evidence that humans may be able to play a more extensive role. For forecasts for which bias is low, Roebber (1998) found that forecasters were nonetheless able to add skill to the model by adjusting their reliance on particular pieces of forecast information according to the meteorological situation. Roebber and Bosart (1996) and Roebber et al. (1996) found that experienced forecasters are better able to recognize those instances when simple forecast strategies do not apply, and showed that local experience is an important contributor to forecast skill. Historically, engaged forecasters who actively appraise their conceptual understanding of the forecast problem have been able to add skill to forecasts, even as the technology improved (e.g., McIntyre 1999; Bosart 2003).

The reasons for this continuing success are known. Although humans have difficulty managing large volumes of data (a trivial task for even simple computers), skilled practitioners are adept at *interpreting* and *evaluating* information, something that remains extraordinarily difficult for automated systems to accomplish. An analogous situation has emerged in medicine, where there is an increasing need to support medical decision making via electronic patient records, automated alerts and reminders, and clinical guidelines (Silverman 1997). Clearly, skilled forecasters are still needed for the foreseeable future to make 1–3-day forecasts because of model deficiencies, underdispersive ensembles, and the simple postprocessing approaches that currently exist at operational centers.

However, decision support systems, no matter how sophisticated, are simply tools and, as such, can lead to adverse effects on forecast skill for disengaged forecasters (Pliske et al. 2004). [See Wickens et al. (1998, 498–504) for a general discussion of human–computer interactions in complex decision environments.] Thus, the considerable effort required to glean value from a high-resolution model in a research setting cannot be transplanted effectively into operations without sufficient forecaster education and training, so that forecasters possess the scientific knowledge required to intelligently interrogate the model data (e.g., Doswell et al. 1981; Doswell 1986; Tennekes 1988, 1992; Brooks et al. 1992; Brooks and Doswell 1993). Rigorous forecaster training needs to be accomplished before people are hired and on shift, and this training needs to be extended throughout forecasters' careers in ways that encourage the continued development and application of forecast techniques that can be applied in local offices, consistent with advances in technology.

Different modes of dissemination should also be considered to overcome transmission bottlenecks. As the above discussion suggests, valuable information can be gained from the complementary use of high-resolution models and ensembles. Regionalization, under controlled conditions in which the necessary training and support is provided, will help maintain the human element in the forecast enterprise, by providing for forecaster input into model validation and by promoting the use of model forecasts as a tool rather than an oracle. Furthermore, each region has its own set of inherent forecast issues that may be more effectively addressed locally with custom high-resolution model setups, mesoscale data assimilation, and ensembles. We recommend that large-domain moderate-resolution forecasts and/or ensembles be generated at a central facility, and disseminated to regional centers for input to high-resolution models and additional ensembles.

As noted previously, the current paucity of observational data at all scales remains a serious constraint to forecasting. In particular, forecasters who use high-

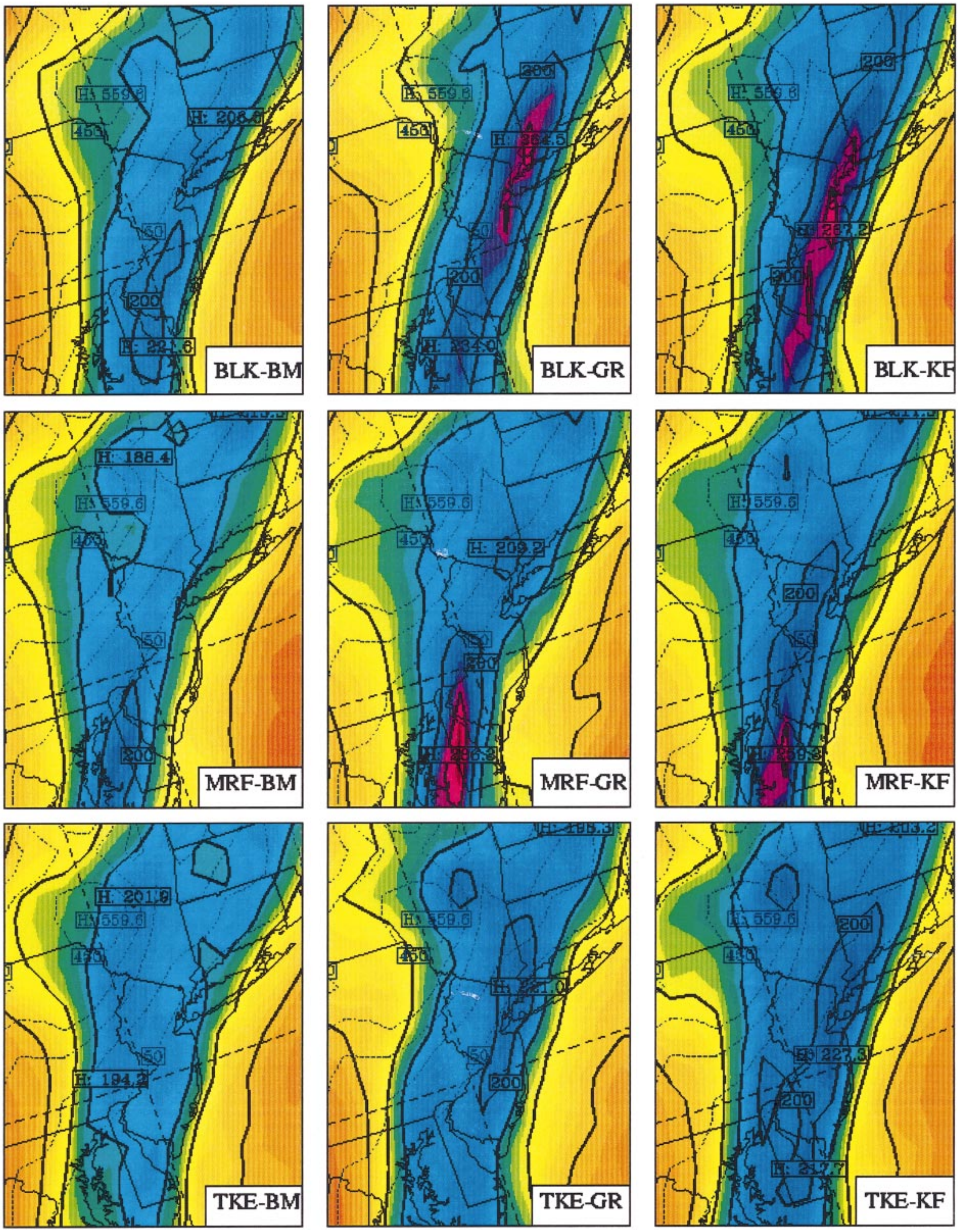

FIG. 2. Nine ensemble members for Tropical Storm Floyd showing the 6–36-h (0600 UTC 16 Sep 1999–1200 UTC 17 Sep 1999) storm total precipitation (solid black lines every 50 mm and color shaded) for each of the nine physics members at 32-km grid spacing. The terrain from the 32-km domain is dashed (every 100 m starting at 50 m). Individual ensemble members are identified with the following notation:

resolution model output without the means to test and verify its output, may be unable to perform hypothesis testing at the required scales, with the likely result being the supplanting of observational diagnosis entirely (e.g., Bosart 2003). In order for efforts at ensemble or high-resolution NWP to provide value, as well as for real-time evaluation of model performance by forecasters, there must be increased collection of observational data at multiple scales.

Because technological progress continues at a rapid pace, some of the constraints on the production of high-resolution guidance are temporary. However, limitations in observations will establish effective limits to forecast resolution, even in the face of continuing improvements in computing and scientific understanding. Ensembles will always be necessary to provide guidance concerning uncertainties imposed by observational and physics limitations. Forecasting experiments (e.g., Kain et al. 2003) should be a routine part of National Weather Service (NWS) operations, as a means of exploring the combined use of high-resolution and ensemble forecast models. Along these lines, postprocessing of data is a seriously underutilized aspect of NWP that, if thoroughly investigated, could lead to substantial forecast improvements. A scientific forecasting service should be maintained and further encouraged by active exploration of model postprocessing techniques at local offices.

Even with these changes in operations, there are several key science problems that continue to limit potential improvements in forecasting and require further research: a lack of a basic understanding of mesoscale predictability, deficient data assimilation, and incomplete or inaccurate representation of atmospheric physics in NWP. For which atmospheric phenomena of forecast interest are we nearing the practical limits to predictability imposed by chaotic dynamics? Which can be expected to yield to further practical and theoretical improvements? Why do the benefits of mesoscale data assimilation using high-resolution observations only extend 6–12 h into the future (possible answers: the domains are too small, the model physics are poor, the forecast range is beyond the limits of predictability)? It is clear that both high resolution and ensembles suffer from crude physical parameterizations, particularly for the planetary boundary layer and microphysical processes (e.g., section 2c). Active investigations in these areas, made possible by robust funding mechanisms that support operational research, must be developed.

Failure to address these limitations can only lead to a convergence of model and human forecast skill at levels well below the optimum. The costs of this failure are unknown, but available evidence suggests that they would be high. Extreme weather is estimated to cost the U.S. economy \$30 billion yr^{-1} [(the National Oceanic and Atmospheric Administration) NOAA 2002]. Impacts from day-to-day weather variations may well be much larger, given that an estimated \$3.8 trillion of the economy (39%) is exposed to weather hazards (Dutton 2002). In contrast, the operation and maintenance of observing systems (excluding satellite data, but including fundamental data such as the radiosonde network, WSR-88D, and surface reports) costs approximately \$90 million yr^{-1} (based on the 2002 NOAA budget). With the addition of approximately \$700 million for the National Environmental Satellite Data and Information Service (NESDIS; which provides for polar-orbiting and geostationary satellites, as well as the development, production, and distribution of products from these satellites), the costs of observations represent less than 3% of that of severe weather and 0.02% of the total U.S. weather exposure. Completely neglecting the issue of public safety, a simple cost–loss calculation suggests that even minimal improvements in our ability to predict, and hence protect, against adverse weather events will be cost effective. A coordinated agenda of improved model formulations, enhanced forecaster training in the scientific use of model guidance in the forecast process, and improved spatial and temporal sampling of the atmosphere should be viewed as an essential effort in a society seeking to use limited funds in the most effective manner possible in support of the national interest.

Acknowledgments. Some of this work was performed during the first author's sabbatical at the NSSL/SPC in December 2000. Further work occurred during the second author's stay at University of Wisconsin—Milwaukee in October 2001. The use of computing time and resources, system administration, and hospitality offered by these organizations is gratefully acknowledged. Use of the MM5 was made possible by the Mesoscale and Microscale Meteorology Division of NCAR. This paper has benefited from discussions with Lance Bosart, Charles Doswell, John Nielsen-Gammon, Jack Kain, Daryl Onton, and Jim Steenburgh. Comments on an earlier version of this manuscript were graciously provided by David Bright, David Novak, Phil Schumacher, Jim Steenburgh, and Steve Weiss. The constructive criticisms of the editor and two anonymous reviewers were important to the final form of this paper. This work was funded in part by NSF Grant OCE-9726679-03 (Roebber), UCAR/COMET Grants SOO-18116 (Roebber) and S01-24224 (Colle), the Office of Naval Research Grant N000014-00-1-0407 (Colle), and NOAA/OAR/NSSL

←

FIG. 2. (*Continued*) XXX-YY, where XXX represents the boundary layer parameterization: BLK = Blackadar, MRF = Medium-Range Forecast model scheme, TKE = Burk and Thompson; and YY represents the convective parameterization scheme: BM = Betts–Miller–Janjic, GR = Grell, KF = Kain–Fritsch.

under NOAA-OU Cooperative Agreement NA17RJ1227 (Schultz).

APPENDIX

Scientific Forecasting

Scientific forecasting consists of three steps: hypothesis formation, hypothesis testing, and prediction (e.g., Doswell 1986; Doswell and Maddox 1986; Hoffman 1991; Pliske et al. 2004). *Hypothesis formation* requires that the forecaster develop a conceptual understanding of the forecast scenario by diagnosing the current state of the atmosphere and making intelligent choices about possible forecast scenarios, which allows the forecaster to develop situational awareness (usually through the accomplishment of specific forecast tasks informed by the details of the forecast situation—the so-called problem of the day). The second step, *hypothesis testing,* requires that the forecaster seek evidence to confirm or refute the hypothesis. This evidence will depend on the specific weather situation being forecast, the associated hypothesis, the tools available to the forecaster, and the availability of observational or modeling data. During this stage, ongoing inspection of data and evaluation of previous short-term forecasts occurs, typically, simultaneous with the issuance of specific forecasts; thus, development and reappraisal of the forecaster's conceptual understanding of the situation is continuous (e.g., Perby 1989). Hypothesis testing is an iterative process through which the eventual outcome of *prediction* is achieved.

Pliske at al. (2004) analyzed the forecast process as conducted by U.S. Air Force and NWS forecasters. They found a variety of forecasting styles: 1) *intuitive forecasters,* who constructed their conceptual understanding on the basis of dynamic, visual images; 2) *rule-based forecasters,* who relied upon extensive knowledge of meteorological rules of thumb, but were informed by an understanding of dynamics; 3) *procedure-based forecasters/mechanics,* who approached forecasting as a procedural task without developing any conceptual understanding, but achieved a rote-based efficiency for forecasting at a specific location; and 4) *disengaged forecasters,* who seemed neither motivated to improve their forecast skill/meteorological knowledge nor displayed proficiency in conducting the forecast task.

Pliske et al. (2004) found that intuitive forecasters were adept at using and integrating information from a variety of sources and thereby detecting patterns in the data during the hypothesis-formation and hypothesis-testing stages. They gradually built their conceptual understanding of the forecast scenario from subelements of the days' weather. Forecasters then critically examined this understanding through evaluation of additional datasets, such as model output and updated observations. Inspection of available data often led these forecasters to recognition-primed decision making. For example, if the ingredients for convection [i.e., moisture, instability, and lift; e.g., Johns and Doswell (1992)] were simultaneously present with favorable wind shear profiles, an intuitive forecaster would recognize the potential for severe storms and/or flash flooding and adjust their hypothesis testing accordingly.

High-resolution model data may provide an additional, powerful means for assessment of a forecaster's conceptual understanding, both early in the development process and later as a means of interpreting the incoming observational data stream. This use of the model for interpretation of observations is made possible to the extent that the verifiable model output mirrors key aspects of the existing (albeit incomplete) observations, since the model provides a four-dimensional dataset at the designated temporal and spatial scales. Further, the high-resolution model output provides insight into possible outcomes that lower-resolution model output cannot directly provide.

As discussed by Brooks and Doswell (1993), however, if a single model forecast goes awry, the forecaster is left with no numerical guidance and must rely only on the available data and their knowledge of past weather events. Unfortunately, the model solutions tend to exhibit the least certainty (e.g., as evidenced by inconsistency from run to run) in the most dynamic situations. Ensembles, by placing the information in a probabilistic format, have the potential to directly address this problem. Ensemble-model output in the hands of engaged forecasters may be valuable in evaluating alternate forecast scenarios, especially in the case of unlikely, but potentially devastating, events. Numerical output describing the probability of such scenarios may assist the forecaster in developing probabilistic forecast products, such as those in use at the Storm Prediction Center (e.g., Leftwich et al. 1998; Kay and Brooks 2000). Whether using high-resolution or ensemble models, engaged forecasters are better able to assess the quality of the model forecasts and incorporate these conclusions into their thinking. Forecasters who incompletely develop or fail to develop a conceptual understanding of the forecast situation, however, will be incapable of taking full advantage of the available NWP.

REFERENCES

Andra, Jr., D. L., E. M. Quoetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy and forecaster expertise on 3 May 1999. *Wea. Forecasting,* **17,** 559–566.

Atallah, E. H., and L. F. Bosart, 2003: The extratropical transition and precipitation distribution of Hurricane Floyd (1999). *Mon. Wea. Rev.,* **131,** 1063–1081.

Atger, F., 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. *Mon. Wea. Rev.,* **131,** 1509–1523.

Baldwin, M. E., S. Lakshmivarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. Preprints, *Ninth Conf. on Mesoscale Processes,* Fort Lauderdale, FL, Amer. Meteor. Soc., 255–258.

Ballentine, R. J., A. J. Stamm, E. F. Chermack, G. P. Byrd, and D. Schleede, 1998: Mesoscale model simulation of the 4–5 January 1995 lake-effect snowstorm. *Wea. Forecasting,* **13,** 893–920.

Bernadet, L. R., L. D. Grasso, J. E. Nachamkin, C. A. Finley, and W. R. Cotton, 2000: Simulating convective events using a high-resolution mesoscale model. *J. Geophys. Res.,* **105,** 14 963–14 982.

Black, T. L., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Bosart, L. F., 1990: Degradation of the North American radiosonde network. *Wea. Forecasting,* **5,** 527–532.

——, 2003: Whither the weather analysis and forecasting process? *Wea. Forecasting,* **18,** 520–529.

Bright, D. R., S. J. Weiss, J. J. Levit, and D. J. Stensrud, 2003: The utility of short-range ensemble forecasts in the real-time prediction of severe convective weather at the Storm Prediction Center. Preprints, *10th Conf. on Mesoscale Processes,* Portland, OR, Amer. Meteor. Soc., 2.8.

Brooks, H. E., and C. A. Doswell III, 1993: New technology and numerical weather prediction—A wasted opportunity? *Weather,* **48,** 173–177.

——, ——, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting,* **7,** 120–133.

Colle, B. A., 2003: Numerical simulations of the extratropical transition of Floyd (1999): Structural evolution and responsible mechanisms for the heavy rainfall over the northeast United States. *Mon. Wea. Rev.,* **131,** 2905–2926.

——, and C. F. Mass, 2000: The 5–9 February 1996 flooding event over the Pacific Northwest: Sensitivity studies and evaluation of the MM5 precipitation forecasts. *Mon. Wea. Rev.,* **128,** 593–617.

——, K. J. Westrick, and C. F. Mass, 1999: Evaluation of MM5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. *Wea. Forecasting,* **14,** 137–154.

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

——, ——, and D. Ovens, 2001: Evaluation of the timing and strength of MM5 and Eta surface trough passages over the eastern Pacific. *Wea. Forecasting,* **16,** 553–572.

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

Doswell, C. A., III, 1986: The human element in weather forecasting. *Natl. Wea. Dig.,* **11,** 6–17.

——, and R. A. Maddox, 1986: The role of diagnosis in weather forecasting. Preprints, *11th Conf. on Weather Forecasting and Analysis,* Kansas City, MO, Amer. Meteor. Soc., 177–182.

——, L. R. Lemon, and R. A. Maddox, 1981: Forecaster training—A review and analysis. *Bull. Amer. Meteor. Soc.,* **62,** 983–988.

Douglas, M. W., and D. J. Stensrud, 1996: Upgrading the North American upper-air observing network: What are the possibilities? *Bull. Amer. Meteor. Soc.,* **77,** 907–924.

Droegemeier, K. K., and Coauthors, 2002: Project CRAFT: A test bed for demonstrating the real time acquisition and archival of WSR-88D Level II data. Preprints, *18th Int. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* Orlando, FL, Amer. Meteor. Soc., 136–139.

Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.,* **121,** 1493–1513.

Dutton, J. A., 2002: Opportunities and priorities in a new era for weather and climate services. *Bull. Amer. Meteor. Soc.,* **83,** 1303–1311.

Elmore, K. L., D. J. Stensrud, and K. C. Crawford, 2002: Ensemble cloud model applications to forecasting thunderstorms. *J. Appl. Meteor.,* **41,** 363–383.

Emanuel, K., and Coauthors, 1995: Report of the First Prospectus Development Team of the U.S. Weather Research Program to NOAA and the NSF. *Bull. Amer. Meteor. Soc.,* **76,** 1194–1208.

Fowle, M. A., and P. J. Roebber, 2003: Short-range (0–48-h) numerical prediction of convective occurrence, mode and location. *Wea. Forecasting,* **18,** 782–794.

Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.,* **79,** 285–299.

Gallus, W. A., 1999: Eta simulations of three extreme precipitation events: Sensitivity to resolution and convective parameterization. *Wea. Forecasting,* **14,** 405–426.

——, 2002: Impact of verification grid-box size on warm-season QPF skill measures. *Wea. Forecasting,* **17,** 1296–1302.

——, and M. Segal, 2001: Impact of improved initialization of mesoscale features on convective system rainfall in 10-km Eta simulations. *Wea. Forecasting,* **16,** 680–696.

Glahn, H. R., 2003: Comments on ''IFPS and the future of the National Weather Service.'' *Wea. Forecasting,* **18,** 1299–1304.

——, and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 138 pp. [Available from NCAR, P. O. Box 3000, Boulder, CO 80307-3000.]

Grimit, E. P., and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting,* **17,** 192–205.

Hall, T., H. E. Brooks, and C. A. Doswell III, 1999: Precipitation forecasting using a neural network. *Wea. Forecasting,* **14,** 338–345.

Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

Harr, P. A., and R. L. Elsberry, 2000: Extratropical transition of tropical cyclones over the western North Pacific. Part I: Evolution of structural characteristics during the transition process. *Mon. Wea. Rev.,* **128,** 2613–2633.

Hart, K. A., W. J. Steenburgh, D. J. Onton, and A. J. Siffert, 2004: An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Wea. Forecasting,* **19,** 200–218.

Heideman, K. F., T. R. Stewart, W. R. Moninger, and P. Reagan-Cirincione, 1993: The Weather Information and Skill Experiment (WISE): The effect of varying levels of information on forecast skill. *Wea. Forecasting,* **8,** 25–36.

Hoffman, R. R., 1991: Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Wea. Forecasting,* **6,** 98–110.

Janjic, Z. I., J. P. Gerrity Jr., and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.,* **129,** 1164–1178.

Jankov, I., and W. A. Gallus, 2004: MCS rainfall forecast accuracy as a function of large-scale forcing. *Wea. Forecasting,* **19,** 428–439.

Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting,* **7,** 588–612.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.,* **84,** 1797–1806.

Kay, M. P., and H. E. Brooks, 2000: Verification of probabilistic severe storm forecasts at the SPC. Preprints, *20th Conf. on Severe Local Storms,* Orlando, FL, Amer. Meteor. Soc., 285–288.

Koizumi, K., 1999: An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network. *Wea. Forecasting,* **14,** 109–118.

Krishnamurti, T. N., and Coauthors, 2001: Real-time multianalysis–

multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.,* **129,** 2861–2883.

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

Langland, R. H., M. A. Shapiro, and R. Gelaro, 2002: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 east coast snowstorm. *Mon. Wea. Rev.,* **130,** 957–974.

Larson, V. E., R. Wood, P. R. Field, J.-C. Golaz, T. H. Vonder Haar, and W. R. Cotton, 2001: Systematic biases in the microphysics and thermodynamics of numerical models that ignore subgrid-scale variability. *J. Atmos. Sci.,* **58,** 1117–1128.

Leftwich, P. W., Jr., J. T. Schaefer, S. J. Weiss, and M. Kay, 1998: Severe convective storm probabilities for local areas in watches issued by the Storm Prediction Center. Preprints, *19th Conf. on Severe Local Storms,* Minneapolis, MN, Amer. Meteor. Soc., 548–551.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—Has its time come? *Quart. J. Roy. Meteor. Soc.,* **116,** 779–798.

Lynn, B. H., A. P. Khain, J. Dudhia, D. Rosenfeld, A. Pokrovsky, and A. Seifert, 2004a: Spectral (bin) microphysics coupled with a mesoscale model (MM5). Part I: Model description and first results. *Mon. Wea. Rev.,* in press.

——, ——, ——, ——, ——, and ——, 2004b: Spectral (bin) microphysics coupled with a mesoscale model (MM5) Part II: Simulation of a CaPe rain event with squall line. *Mon. Wea. Rev.,* in press.

Manobianco, J., and P. A. Nutter, 1999: Evaluation of the 29-km Eta model. Part II: Subjective verification over Florida. *Wea. Forecasting,* **14,** 18–37.

Marshall, C. H., K. C. Crawford, K. E. Mitchell, and D. J. Stensrud, 2003: The impact of land surface physics in the operational NCEP Eta Model on simulating the diurnal cycle: Evaluation and testing using Oklahoma Mesonet data. *Wea. Forecasting,* **18,** 748–768.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.,* **127,** 2095–2116.

Mass, C. F., 2003a: IFPS and the future of the National Weather Service. *Wea. Forecasting,* **18,** 75–79.

——, 2003b: Reply. *Wea. Forecasting,* **18,** 1305–1306.

——, D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

McBride, J. L., and E. Ebert, 2000: Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Wea. Forecasting,* **15,** 103–121.

McIntyre, M. E., 1999: Numerical weather prediction: A vision of the future, updated still further. *The Life Cycles of Extratropical Cyclones,* M. A. Shapiro and S. Grønås, Eds., Amer. Meteor. Soc., 337–355.

McMurdie, L., and C. Mass, 2004: Major numerical forecast failures over the Northeast Pacific. *Wea. Forecasting,* **19,** 338–356.

Mesinger, F., 1996: Forecasting cold surges east of the Rocky Mountains. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 68–69.

Molteni, F., R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. I: Definition of representative members and global model experiments. *Quart. J. Roy. Meteor. Soc.,* **127,** 2069–2094.

Mullen, S. L., and R. Buizza, 2002: The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Wea. Forecasting,* **17,** 173–191.

National Research Council, 1998: *The Atmospheric Sciences Entering the Twenty-First Century.* National Academy Press, 364 pp.

Nielsen-Gammon, J. W., and J. Strack, 2000: Model resolution dependence of simulations of extreme rainfall rates. Preprints, *10th PSU/NCAR Mesoscale Model Users Workshop,* Boulder, CO, PSU/NCAR, 110–111. [Available from MMM Division, NCAR, P. O. Box 3000, Boulder, CO 80303.]

NOAA, 2002: NOAA Economic Statistics. U.S. Department of Commerce, Office of Policy and Strategic Planning, 23 pp. [Available online at http://www.publicaffairs.noaa.gov/worldsummit/pdfs/economicstats.pdf.]

Orrell, D., L. Smith, J. Barkmeijer, and T. N. Palmer, 2001: Model error in weather forecasting. *Nonlinear Processes Geophys.,* **8,** 357–371.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.,* **128,** 747–774.

Perby, M.-L., 1989: Computerization and skill in local weather forecasting. *Knowledge, Skill, and Artificial Intelligence,* B. Göranson and I. Josefson, Eds., Springer-Verlag, 39–52.

Petersen, R. A., and J. D. Stackpole, 1989: Overview of the NMC production suite. *Wea. Forecasting,* **4,** 313–322.

Peyraud, L., 2001: Radiation fog forecasting using a 1-dimensional model. M.S. thesis, Dept. of Atmospheric Sciences, Texas A&M University, 82 pp. [Available from Dept. of Atmospheric Sciences, Texas A&M University, 3150 TAMU, College Station, TX 77843-3150.]

Pliske, R., B. Crandall, and G. Klein, 2004: Competence in weather forecasting. *Psychological Investigations of Competent Decision Making,* J. Shanteau, P. Johnson, and K. Smith, Eds., Cambridge University Press, in press.

Roebber, P. J., 1998: The regime dependence of degree day forecast technique, skill, and value. *Wea. Forecasting,* **13,** 783–794.

——, and L. F. Bosart, 1996: The contributions of education and experience to forecast skill. *Wea. Forecasting,* **11,** 21–40.

——, and M. G. Gehring, 2000: Real-time prediction of the lake breeze on the western shore of Lake Michigan. *Wea. Forecasting,* **15,** 298–312.

——, and J. R. Gyakum, 2003: Orographic influences on the mesoscale structure of the 1998 ice storm. *Mon. Wea. Rev.,* **131,** 27–50.

——, L. F. Bosart, and G. J. Forbes, 1996: Does distance from the forecast site affect skill? *Wea. Forecasting,* **11,** 582–589.

——, D. M. Schultz, and R. Romero, 2002: Synoptic regulation of the 3 May 1999 tornado outbreak. *Wea. Forecasting,* **17,** 399–429.

Schlatter, T. W., and S. J. Lord, 1997: The scientific basis for NAOS. Preprints, *First Symp. on Integrated Observing Systems,* Long Beach, CA, Amer. Meteor. Soc., 3–5.

Schultz, D. M., and C. A. Doswell III, 2000: Analyzing and forecasting Rocky Mountain lee cyclogenesis often associated with strong winds. *Wea. Forecasting,* **15,** 152–173.

Schumacher, P., 2003: An example of forecasting mesoscale bands in an operational forecasting environment. Preprints, *10th Conf. on Mesoscale Processes,* Portland, OR, Amer. Meteor. Soc., p1.11.

Schwartz, G., 1980: Death of the NWS forecaster. *Bull. Amer. Meteor. Soc.,* **61,** 36–37.

Schwerdtfeger, W., 1981: Comments on Tor Bergeron's contributions to synoptic meteorology. *Pure Appl. Geophys.,* **119,** 501–509.

Silverman, B. G., 1997: The initial failure of artificial inteligence in medicine (AIM), the rise of the grand challenge, and a new role for AIM. Panel presentation for the National Science Foundation Workshop on Human Centered Systems. [Available online at http://www.ifp.uiuc.edu/nabhcs/abstracts/silverman.html.]

Steenburgh, W. J., and D. J. Onton, 2001: Multiscale analysis of the 7 December 1998 Great Salt Lake–effect snowstorm. *Mon. Wea. Rev.,* **129,** 1296–1317.

Stensrud, D. J., H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433–446.

——, J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077–2107.

Stewart, T. R., K. F. Heideman, W. R. Moninger, and P. Reagan-Cirincione, 1992: Effects of improved information on the components of skill in weather forecasting. *Org. Behavior Human Decision Processes,* **53,** 107–134.

Tennekes, H., 1988: The outlook: Scattered showers. *Bull. Amer. Meteor. Soc.,* **69,** 368–372.

——, 1992: Karl Popper and the accountability of numerical weather prediction. *Weather,* **47,** 343–346.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting,* **8,** 379–400.

Uccellini, L. W., P. J. Kocin, and J. M. Sienkiewicz, 1999: Advances in forecasting extratropical cyclogenesis at the National Meteorological Center. *The Life Cycles of Extratropical Cyclones,* M. A. Shapiro and S. Grønås, Eds., Amer. Meteor. Soc., 317–336.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved model output statistics forecasts through model consensus. *Bull. Amer. Meteor. Soc.,* **76,** 1157–1164.

Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.,* **129,** 729–747.

Warner, T. T., and H.-M. Hsu, 2000: Nested-model simulation of moist convection: The impact of coarse-grid parameterized convection on fine-grid resolved convection. *Mon. Wea. Rev.,* **128,** 2211–2231.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.,* **125,** 527–548.

Weygandt, S. S., and N. L. Seaman, 1994: Quantification of predictive skill for mesoscale and synoptic-scale meteorological features as a function of horizontal grid resolution. *Mon. Wea. Rev.,* **122,** 57–71.

——, A. Shapiro, and K. K. Droegemeier, 2002: Retrieval of model initial fields from single-Doppler observations of a supercell thunderstorm. Part II: Thermodynamic retrieval and numerical prediction. *Mon. Wea. Rev.,* **130,** 454–476.

Wickens, C. D., S. E. Gordon, and Y. Liu, 1998: *An Introduction to Human Factors in Engineering.* Addison-Wesley Educational Publishers Inc., 636 pp.

Wright, P., 1974: The harassed decision maker: Time pressures, distractions and the use of evidence. *J. Appl. Psychol.,* **59,** 555–561.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the ''surprise'' snowstorm of 24–25 January 2000. *Mon. Wea. Rev.,* **130,** 1617–1632.

Zupanski, D., D. F. Parrish, E. Rogers, and G. DiMego, 2002: Four-dimensional variational data assimilation for the blizzard of 2000. *Mon. Wea. Rev.,* **130,** 1967–1988.