1. Introduction

A workshop on ensemble forecasting in the short to medium ranges (0–14 days forecast lead time) was held at the National Center for Atmospheric Research in Boulder, Colorado, 9–11 September 1999. Approximately 45 people attended this workshop, with approximately a quarter joining us from outside the United States. The purpose of the workshop was to discuss the current state of knowledge of ensemble forecasting, to define the most important research problems for the next few years, and to seek common evaluation methods and tools. The sessions in this workshop were organized around three general topics: 1) use of ensemble forecasts for data assimilation, 2) issues related to model error in ensemble forecasts, and 3) the use, utility, and interpretation of ensemble forecasts.

In this meeting summary, we first provide some background on ensemble forecasting. We also compare the current state of our knowledge to what was known at the last U.S. ensemble forecasting workshop in 1994 (Brooks et al. 1995). Next, we provide summaries and recommendations from each of the three workshop sessions and end with a brief conclusion.

2. Background and recent progress in ensemble forecasting

Ensemble forecasting (EF) has been embraced as a practical way of estimating the uncertainty of a weather forecast. Since Lorenz (1963, 1969) it has been recognized that perfect numerical weather forecasts will always be unattainable; even the smallest of errors in the initial conditions will grow inexorably, eventually rendering any single deterministic forecast useless. Rather than pinning unrealistic hopes upon the accuracy of a single numerical forecast, EF adopts an alternative approach: generate multiple, individual numerical forecasts from different initial conditions and/or different numerical model configurations (Leith 1974). Probabilistic forecasts of the weather may then be generated from the relative frequencies of events in the en-
semble. Medium-range EFs have been produced operationally in the United States and Europe since late 1992 (Toth and Kalnay 1993, 1997; Palmer et al. 1993; Molteni et al. 1996). At the Canadian Meteorological Centre (CMC), EFs have been produced operationally since January 1996 (Houtekamer et al. 1996a,b; Houtekamer and Lefaitvre 1997). Each of the centers produces the ensembles using different forecast models and different ensemble construction techniques. EFs are also now produced operationally at several other centers around the world as well (e.g., Rennick 1995; Kobayashi et al. 1996).

Encouraging results from medium-range EFs motivated the previous workshop on ensemble forecasting, which focused on the potential utility of short-range ensemble forecasts (SREFs; Brooks et al. 1995). It was believed that SREFs might provide useful information for short-range forecast problems such as severe storms forecasting (Brooks et al. 1992) and precipitation forecasting. At that workshop, issues of how to generate ensemble forecasts were considered, as well as considerations of model error and ways of dealing with the data overload of the multiple weather forecasts. It was recognized that the problems for generating useful SREFs were potentially much more difficult than for medium-range EFs; many systematic errors corrupt our current mesoscale numerical weather forecasts, owing to insufficient model resolution, use of physical parameterizations, insufficient knowledge of the land surface condition, and other such problems. As a result of this workshop, the National Centers for Environmental Prediction (NCEP) launched a pilot project to generate a small ensemble of SREFs using a reduced-resolution version of theEta model (Black 1994) and the Regional Spectral Model (Juang and Kanamitsu 1994). These EFs were initialized with both interpolated bred initial conditions from NCEP's medium-range EF and from a variety of in-house analyses.

Since the 1994 workshop much has been learned about EFs and SREFs. Results from the pilot SREF project were described in Hamill and Colucci (1997, 1998) and Stensrud et al. (1999). Based on these and other recent results, NCEP plans to implement a semioperational SREF starting in 2000. Other work on EFs since the 1994 workshop includes, for example, case studies of intense cyclogenesis (Du et al. 1997; Leslie and Speer 1998; Hamill 1998; Mullen et al. 1999) and blocking (Colucci and Baumhefner 1998); the study of the performance of EFs, for precipitation forecasting (Hamill and Colucci 1997, 1998; Du et al. 1997; Eckel and Walters 1998; Buizza et al. 1999a; Mullen and Buizza 2000); a study of the benefit of postprocessing EFs (Eckel and Walters 1998); a synoptic evaluation of the NCEP medium-range ensemble (Toth et al. 1997); exploration of issues related to the choice of perturbation methodologies using perfect models (Houtekamer and Derome 1995; Anderson 1997; Hamill et al. 2000); a comparison of the relative effects of model and initial condition errors in the presence of convection (Stensrud et al. 2000); effects of increasing the ensemble size (Buizza and Palmer 1998) and the resolution of member forecasts (Buizza et al. 1998); the effects of domain size and lateral boundaries (Du and Tracton 1999); methods for evaluating EFs (Anderson 1996; Smith and Gilmour 1999; Wilson et al. 1999); examinations of spread-skill relationships in ensembles (Buizza 1997; Whitaker and Loughe 1998); and examinations of the potential utility of ensembles from multiple models and/or multiple initial conditions (Krishnamurti et al. 1999; Harrison et al. 1999; Evans et al. 2000; Hou et al. 2000, Ziehmann 2000; Richardson 2000b, manuscript submitted to Quart. J. Roy. Meteor. Soc.). Ehrendorfer (1997) and Palmer (2000) provide nice reviews of EF concepts.

Research in the use of ensemble forecasts for improving data assimilation has also blossomed. A crucial part of any data assimilation methodology is the specification of error statistics for the first-guess or "background" forecast. These statistics determine how much to weight the background relative to the observations and how to spread the influence of the observations away from the actual observation location. The accuracy of analyses and subsequent forecasts can potentially be improved greatly if background error covariances are better estimated. Older data assimilation methods such as optimum interpolation (Gandin 1963; Schlatter 1975; Lorenc 1981) and three-dimensional variational analysis (3D-Var; Lorenc 1986; Parrish and Derber 1992) use simple statistics for describing the background errors, which may not vary in time or space. Recent results suggest that it may be possible to generate more accurate spatially and temporally varying background error statistics from a set of EFs. Articles by Evensen (1994), Houtekamer and Mitchell (1998), Burgers et al. (1998), van Leeuwen (1999), Mitchell and Houtekamer (2000), and Hamill and Snyder (2000) discuss the use of an ensemble of forecasts using a technique called the "ensemble Kalman Filter" (EnKF). The EnKF is a special case of the nonlinear...
filter discussed by Anderson and Anderson (1999). Also, the use of ensembles in a reduced-rank extended Kalman filter is discussed by Fisher and Courtier (1995) and Fisher (1998), the use of bred mode information for improving analyses is discussed in Pu et al. (1997), and the use of ensemble forecast statistics for specifying improved stationary background error statistics to four-dimensional variational analysis (4D-Var) is described in Buizza and Palmer (1999).

Despite many advances in our understanding of how to use and construct ensembles, questions abound. We judged the three most important issues to be 1) how or how best to use ensemble information to improve data assimilation strategies, 2) how to address model errors in EFs, and 3) how to appropriately use and interpret the voluminous information from ensembles.

The workshop was organized into three sessions, each dealing with one of these issues. In each session, there was a set of invited, longer talks, a larger number of shorter presentations, extensive discussion, and a final working group tasked with summarizing the current state of the art and recommending areas requiring further research. A report from each session follows.

3. Session 1: Ensemble forecasting and data assimilation

A common theme ran through the presentation and discussion in this session: ensemble forecasting and data assimilation are two aspects of the same problem, namely, describing the evolution of the probability distribution for the atmospheric state given available observations and a forecast model. The presentations here reinforced the supposition that properly constructed ensembles may generate probabilistic information in the very short range that may be used to estimate background error statistics for data assimilation (DA) schemes, consequently improving the accuracy of analyses. Similarly, the generation of an ensemble of initial conditions for purposes of data assimilation should incorporate probabilistic information on analysis errors provided by the data assimilation scheme, which are affected by dynamically constrained errors and model errors (from the background) and by random errors (from assimilating imperfect observations). This general approach to the construction of ensemble initial conditions differs from those used in present operational ensemble systems at the European Centre for Medium-Range Weather Forecasts and NCEP. There, the ensemble of initial conditions are designed for forecast applications and project upon features that will grow or have grown rapidly, respectively.

a. Summary of presentations

An overview of the use of ensembles in data assimilation, with emphasis on the EnKF, was presented by P. Houtekamer (CMC). The EnKF is related to the Kalman filter (Kalman 1960), which provides the optimal estimate of the state of a linear dynamical system under the assumption that observational and background error statistics are precisely known and are Gaussian. The EnKF, however, differs from the Kalman filter in that the error covariance is estimated from an ensemble of short-range, nonlinear forecasts; at analysis times, each member is then updated in such a way that the ensemble perturbations approximate a random sample from the analysis error distribution. The effectiveness of the EnKF, even for small ensembles (~10 to 100 members) has been demonstrated in simple models, but it remains untested in operational numerical weather prediction (NWP).

Houtekamer discussed three practical difficulties related to operational implementation of the EnKF:

1) Rank deficiency. Since feasible ensembles in NWP are composed of far fewer members than the degrees of freedom in the forecast model, the ensemble perturbations cannot span the space of model solutions, and their sample covariance matrix is rank deficient. The resulting analysis corrects the background only in the subspace spanned by the ensemble members.

2) Sampling errors. Covariance estimates from a finite sample are subject to sampling errors that decrease only slowly with the size n of the ensemble (as n^{-1/2}; Casella and Berger 1990). Characteristic sampling errors include spurious correlations between widely separated locations and the overestimation of the leading eigenvalues of the covariance matrix.

3) Model error. Estimating background covariances solely from an ensemble of forecasts generated by the same imperfect model ignores the contribution of the error in the forecast model to the uncertainty of the background. Since a key source of model error is the omission or parameterization of unresolved scales, this problem can be expected to be worst at the smallest resolved scales and to lead to
a systematic underestimation of uncertainty at the small scales by the ensemble.

Houtekamer also emphasized that the use of the full nonlinear forecast model and forward operators in the covariance calculations both simplifies the scheme and appears to make it more robust, even though the Kalman filter is formally applicable only for linear dynamics and forward operators.

Presumably, however, there is a point at which a system’s dynamics become sufficiently nonlinear, and the distributions of interest sufficiently non-Gaussian, that a more general approach is required. J. Anderson described how one might build an EF/DA scheme in this regime (Anderson and Anderson 1999). As it turns out, the framework remains similar to the EnKF. An ensemble of short-term forecasts is used to estimate the prior distribution of the atmospheric state; when observations are available, this estimated distribution is updated given the new observations by Bayes’ rule. Finally, a new ensemble of initial conditions is generated as a random sample from the updated distribution. Unlike the EnKF, where all distributions are assumed Gaussian and the ensemble simply provides an estimate of the background covariances, the distributions in this general case are approximated through nonparametric density-estimation techniques (Silverman 1986), which make minimal assumptions on the form of the distribution. Another presentation by M. Berliner further discussed the statistical concepts behind this approach.

Shorter presentations were made on ensemble Kalman filtering approaches (T. Hamill, J. Whitaker, C. Bishop, J. Hansen), on the improvement of 3D-Var using information from bred modes (D. Barker), on singular vectors (R. Gelaro, J. Ahlquist), and on the limits of linearity assumptions in the construction of ensemble perturbations (I. Gilmour).

b. Working group on ensemble forecasting and data assimilation

The working group began with a discussion of the fundamental problems posed by schemes that combine EF and DA. A successful ensemble-based strategy should provide accurate estimates of background error statistics despite imperfect forecast models, despite the imperfect knowledge of the errors in the observations and their relation to the forecast variables, and despite the requirement to use an ensemble whose size is small compared to the degrees of freedom in the forecast model. Opinions varied widely on the feasibility of this enterprise and on which aspects were, in fact, the most problematic. In the end, there was agreement that experiments in simple models had suggested that schemes combining EF and DA were both feasible and useful, but that success of such schemes in more realistic environments such as in operational NWP remained uncertain.

The working group identified three areas where further research could pave the way for tests of such schemes with operational NWP models.

1) Comparing sampling techniques for estimating the background covariance matrix. Two main approaches have been proposed. The first directly applies Monte Carlo techniques to generate an ensemble that is (approximately) a random sample from the distribution at the analysis time, that is, the EnKF. The second approach seeks to obtain more accurate estimates with fewer members by populating the ensemble with those initial perturbations that will evolve into the leading eigenvectors of the background error covariance matrix (Ehrendorfer and Tribbia 1997; Barkmeijer et al. 1998; for a related method, see Pham et al. 1998). Both of these techniques use the same information (estimates of the observation error covariances and a previous ensemble-based estimate of the background error covariances). Experimentation is still required to determine the relative merits of each approach.

2) Dealing with small sample sizes. As discussed by Houtekamer in his presentation, small sample sizes give rise to many of the known difficulties of ensemble-based data assimilation. Several techniques for dealing with many of these difficulties are known. For example, one can ameliorate both rank deficiency and spurious long-range correlations by “localizing” the covariances, either implicitly by excluding distant observations when calculating the analysis at a point, or explicitly by multiplying the covariances by a decreasing function of distance. However, the relative importance of the known difficulties and the efficacy of these proposed solutions is unclear. There is the potential for further problems, such as a lack of balance in the ensemble perturbations, to arise as ensemble-based assimilation schemes are tested in more complex models.

3) Evaluating different methods. At present, several methods for EF/DA strategies have been proposed and tested in simple models. An important next
4. Session 2: Model errors and ensemble forecasts

When ensemble forecasting was first implemented at NCEP (Toth and Kalnay 1993, 1997) and ECMWF (Palmer et al. 1993; Molteni et al. 1996), the approach was to assess forecast uncertainty related to growth of errors in the initial conditions due to large-scale chaotic dynamics. However, forecast uncertainty also arises because imperfect numerical models are used to predict the behavior of the atmosphere. Ensemble forecast systems that simulate uncertainties due to both initial condition and model errors (e.g., Houtekamer et al. 1996a, b; Houtekamer and Lefèvre 1997) may improve the ensemble, providing a more realistic spread of forecast solutions. Hence, there has been much recent interest in designing new techniques for addressing model uncertainty.

Part of model error can be classified as systematic and another part as random or stochastic. Systematic errors are those that can be reproduced if the model is run many times over similar cases; these errors are commonly referred to as “model bias.” Systematic errors are typically a consequence of model formation, such as inadequate parameterization of certain subgrid scale processes. In principle, if systematic errors are known, model forecasts can be corrected (e.g., Dee and Da Silva 1998); in practice, many of the errors may be conditional, dependent upon the occurrence of convection or other processes, making them hard to estimate with finite samples.

Stochastic errors are not reproducible; they arise at each integration time step due to numerical inaccuracies, the finite truncation at some arbitrary scale from the grid spacing, and other inaccuracies that act randomly. Stochastic errors, just like the initial errors, turn in time into the direction of fastest growing perturbation directions, increasing errors associated with atmospheric instabilities (Toth and Kalnay 1997).

a. Summary of presentations

In recent years, different groups developed various methods to account for model related uncertainty in ensemble forecasting. In his presentation, P. Houtekamer described the technique used at the CMC (Houtekamer et al. 1996a, b; Houtekamer and Lefèvre 1997), whereby several versions of an NWP model are developed and used in parallel with each other. These versions possibly differ from each other in horizontal resolution, treatment of orography, convection, radiation parameterization, etc. For each ensemble model integration started with unique and slightly different initial conditions, a different model version is used. The goal is to capture systematic differences or errors in model forecasts, though, as Houtekamer pointed out, the real atmospheric solution still differs more from the ensemble members than the individual forecasts differ from each other.

Next, R. Buizza described the approach recently implemented at ECMWF (Buizza et al. 1999b). There, after each time step within a model integration, stochastic multiplicative noise (described below) is applied to the net parameterized tendencies using a number chosen randomly in the [0.5, 1.5] interval. The goal is to represent stochastic errors in the parameterization of subgrid-scale processes.

P. Sardeshmukh emphasized that all current physical parameterization schemes return only the ex-
pected value of the subgrid-scale feedback on the resolved tendencies, not the full distribution (Fritsch et al. 1998). This can contribute to an error in the ensemble mean as well as the forecast spread. Stochastic noise can be introduced into ensemble parameterizations to sample the distribution of subgrid-scale feedbacks. This noise is typically introduced as additive noise or multiplicative noise. Additive noise is a separate additional noise term in the prognostic equations, while multiplicative noise is noise that is multiplied with an existing time tendency term(s). Special care must be exercised when using stochastic noise, especially multiplicative noise, as its use may strongly change the ensemble mean (Sardeshmukh et al. 2000).

L. Smith discussed whether the stated goals of ensemble forecasting are appropriate ones, given theoretical considerations about model errors (Smith 1999). He argued that because we do not and will never have a perfect model of the atmosphere, it is impossible to design a reliable, or “accountable” forecast system. Even if the initial uncertainty could be perfectly known and sampled in a statistical sense (which is not the case), without a perfect model, it is impossible to carry forward in time the initial uncertainty in a perfectly consistent manner. Smith then proposed an alternative criterion for evaluating the usefulness of ensembles: ensuring that the ensemble is constructed in such a manner that at least one member follows (or “shadows”) the evolution of the real atmosphere (Gilmour and Smith 1997). A lively discussion followed as to whether the enterprise of ensemble forecasting truly was in the dire straits suggested by the talk.

Shorter presentations were made on a variety of topics: use of different convective parameterizations in SREFs (D. Stensrud), use of ensembles for hurricane track prediction (M. Ramamurthy), operational experimentation at the National Centers for Environmental Prediction (J. Du and I. Szunyogh), and multimodel EFs (C. Ziehmann and G. Pellerin).

b. Working group on model errors

Discussion started with the question, what is the main problem caused by the use of imperfect models in ensemble forecasting? Is it that the overall spread of the ensembles is too small (presumably because stochastic model errors are not accounted for), or is it that the spread in general would be adequate but certain parts of the atmospheric attractor are not covered by the model due to systematic errors? What are the relative contributions of each? No clear answers emerged to these questions. This underscores the need for more quantitative studies analyzing the role of model errors.

O. Talagrand suggested that 4D-Var offers a tool for assessing the role of model errors in short-range forecasts. The inability of the model to properly fit observational data within the range of observational errors (i.e., the extent of the misfit) can be a direct measure of model errors. Another obvious measure is a comparison between the spread of an ensemble and the ensemble mean forecast error. Ideally, these two should be similar. The deficiency in spread may thus be used as a measure of the extent of model-related uncertainties. The extent of the problem can also be diagnosed using rank histograms (also known as “Talagrand diagrams;” Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997) and their extension to multivariate domains by the minimal spanning tree method (Smith 1999).

Next, the theoretical issues previously raised by L. Smith were considered. Participants agreed that despite reservations about the ability of operational ensembles to meet strict theoretical measures of consistency, EFs still can provide valuable probabilistic information. Whatever method(s) can further contribute to the practical goal of improving probabilistic forecasts is worthy of consideration, including statistical postprocessing.

The relative merits of the CMC and ECMWF approaches to model uncertainty were discussed next. Each approach was judged useful, but when used independently, they were noted to be able to simulate only systematic or stochastic errors, but not both. Some hybrid of the different methods that can address both issues is clearly desirable. This is an area where further research is strongly recommended.

Because different forecast systems have different biases, a multimodel, multianalysis ensemble may have appeal (e.g., Harrison et al. 1999; Evans et al. 2000; Hou et al. 2000; Richardson 2000b, manuscript submitted to Quat. J. Roy. Meteor. Soc.). Here, perturbations may be centered about different control forecasts and/or use different forecast models. A simpler approach is to form a smaller ensemble from different control forecasts (Krishnamurti et al. 1999; Ziehmann 2000). However, the number of control forecasts that can be combined is limited, and the prediction of rare events from a small ensemble of control forecasts is especially problematic (the rarer the event, the more ensemble members needed to accurately assess the probability of that event). The group
agreed that combining ensembles generated operationally at different centers, preferably after their bias have been removed, shows promise as an approach to ensemble forecasting in the short term.

The closing discussion of the working group focused on a question raised by Z. Toth. In our efforts to ameliorate the problem related to accounting for model uncertainties, shall we try to build, maintain, and develop a host of models that can work well in different situations, ensuring that at least some solutions may have a chance of verifying well? Or, alternatively, shall we attempt to build one model that can possibly encompass all model versions by randomly varying the structure, parameters, and/or stochastic noise within the physics packages? In other words, shall we try to maintain different model versions that each work best under various conditions, or instead, try to build one model that is inclusive of all versions? No consensus was developed, but it was noted that within the United States, the current development of a single regional forecast modeling system (Dudhia et al. 1998) may offer a unique opportunity for the modeling community to address the issue of model uncertainties within the framework of a single model.

Participants at the model errors working group were J. Ahlquist, J. Du, D. Orrell, G. Pellerin, L. Smith, I. Szunyogh, O. Talagrand, and Z. Toth (chair).

5. Session 3: Use, utility, and interpretation of ensemble forecasts

a. Summary of presentations

First, O. Talagrand provided an overview of the verification of EFs. Ideally, a probabilistic prediction ought to have high reliability (i.e., exhibit low conditional bias for each issued forecast probability) and high resolution (ability of different forecasts to distinguish between different events; if the model is reliable, resolution is related to sharpness). Because qualities of probability distributions such as reliability are being evaluated, it is impossible to objectively assess the quality of an individual ensemble forecast; hence, EF systems must be verified over many cases. Talagrand explained how common scoring metrics (Brier Score, Ranked Probability Score, Relative Operating Characteristics curves, rank histograms, etc.) are contaminated by at least three sources of noise: improper estimates of probabilities from small-sized ensembles, insufficient variety and number of cases in the forecast evaluation, and imperfect observations.

D. Richardson then discussed the potential economic value of EFs (Richardson 2000a). Proper evaluation of the benefit of a forecast system to a particular user should involve not only the intrinsic skill of the forecasts, but also knowledge of the weather-sensitivity and decision-making process of the end user. Reliance on skill measures alone may give a misleading impression of forecast value. To illustrate the concept, he considered the impact of ensemble size on forecast value using a simple decision model (Murphy 1994; Katz and Murphy 1997) and output from the ECMWF Ensemble Prediction System (EPS). Probability forecasts derived from the EPS have greater benefit than a deterministic forecast produced by the same model, and for many users they can have more value than a shorter-range deterministic forecast by the same model or a deterministic forecast from a higher resolution model. The additional information in the EPS, which reflects only the uncertainty in the initial conditions, provides a benefit to users equivalent to many years development of the forecast model and assimilation system. While the difference in skill between 10 and 50 members appears relatively small, the larger ensemble size can yield substantial benefit to a range of users. It also appears that an increase in ensemble size to beyond 51 (the current size for the EPS) can provide additional value, especially for extreme (and unlikely) events.

H. Brooks closed the invited presentations with a discussion of the problems in designing ensemble forecasts for mesoscale weather prediction. Brooks started by noting that computing power has now reached the point where it is technologically feasible to run SREFs with mesoscale models (grid spacing of 30 km or less) in real time on local workstations. The construction of a mesoscale EF system with a limited-area model brings with it several basic questions: how do we even create such an ensemble? What can (and should) we infer about the weather from such an ensemble? How do we provide the output in a timely manner so that decision makers and forecasters can use it? And how do we evaluate the ensemble? The answers to these questions may not be the same as they are for longer-range ensembles, where the lead time and response time is greater. Many fields for which predictive information is desired are poorly observed at the appropriate scales; thus, the errors in mesoscale features can be close to saturated in initial analyses, and verification of them is a very uncertain proposition. Moreover, mesoscale weather phenomena (e.g., severe thunderstorms) can be strongly nonlinear and intermittent:
they quickly appear, amplify, then dissipate within the forecast period. Finally, the model error strongly depends on the phenomenon being forecast and likely becomes more important than initial data errors after a very short time (tens of minutes to a few hours) into the forecast.

Shorter presentations were given on SREFs in the United States (S. Tracton), SREFs in Europe (K. Mylne), postprocessing of EFs (F. Eckel), and the economic value of EFs (Z. Toth).

b. Working group on the use and interpretation of ensemble forecasts

The group first discussed the need for researchers, forecasters, and end users to understand the benefits and limitations of ensemble forecasting. In particular, numerical experimentation suggests there may be a time limit to the ultimate predictability of the atmosphere (Lorenz 1969, 1982). This limit is presumed to vary with the forecast variable in question (e.g., 500-hPa geopotential heights are predictable for longer than cloud cover), the scale of the phenomenon (e.g., Van den Dool and Saha 1990), and the spatial and temporal averaging that are performed (Islam et al. 1993; Vannitsem and Nicolis 1998). Ensemble techniques probably cannot change this limit but can, through ensemble averaging, improved initial conditions, and estimation of case-dependent forecast uncertainty, bring the level of forecast performance somewhat closer to this limit.

Predictability limits are uncertain because they can only be estimated from imperfect numerical models. These estimates are strongly dependent on the dispersion properties of the chosen model and on how close the modeled variance is to the observed variance. If a model lacks variance for a particular scale of motion or cannot even resolve it, energy cascades across the scales will be treated improperly, and the predictability estimate can be expected to be unduly optimistic. In the future, it would be particularly useful to perform new predictability experiments using models with smaller grid spacings so a larger part of the spectrum of atmospheric motions and their interactions can be well resolved. Further research into how predictability estimates change with the scale of the phenomenon is also suggested.

Verification and diagnosis of ensemble forecasts were next discussed. A single verification score is generally inadequate for evaluating all of the desired information about the performance of an analysis/forecast. Each measure provides unique information on system performance. For that reason, a suite of verification measures, appropriate for the evaluation of probabilistic forecasts, should be used.

The group agreed for the need to establish a generally accepted, standardized suite of verification scores and diagrams to evaluate ensemble systems. The group’s consensus was that, at a minimum, the following metrics should be used.

1) Probabilistic skill score measures such as the Brier Score, Brier Skill Score, Ranked Probability Score, and/or Ranked Probability Skill Score (Wilks 1995). These scores can provide an overall, single-number metric for judging the quality of probabilistic forecasts. Their very simplicity also prohibits them from being very informative about the nature of probabilistic forecast errors (Murphy and Winkler 1987; Murphy 1991). However, the Brier score can be decomposed into reliability, resolution, and uncertainty terms (Murphy 1973). A similar decomposition for a continuous ranked probability score was proposed recently (Hersbach 2000).

2) Reliability diagrams (Wilks 1995), plotted together with a decomposition of the Brier score and information on the distribution of forecasts issued (the sharpness). Reliability diagrams can provide information on conditional biases of ensemble forecasts. However, as noted in Wilks (1995), they can be noisy and uninformative unless populated over a large set of cases. Recently, Hamill (1997) demonstrated a multcategory reliability diagram that is less sensitive to the number of cases.

3) The Relative Operating Characteristic (ROC) (Swets 1973; Mason 1982; Stanski et al. 1989; Atger 1999). The ROC curve graphs probabilities of incorrect null and alternative hypotheses as each sorted ensemble member is used as a decision-making threshold. The ROC curve is based on stratification by observations; it is independent of reliability and instead provides a measure of resolution. It is particularly valuable for comparing the performance of ensemble systems against single deterministic forecasts at higher resolution, and the more general resource issue of ensemble size/configuration versus model resolution. Moreover, potential economic value for the simple decision model discussed by Richardson (2000a) is uniquely determined from ROC curves.

4) Rank histograms (Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997), as well as their extension to higher dimensions by the minimal spanning tree (Smith 1999). These diagnose
the ability of the ensemble to sample from the correct probability distribution. Model bias and under- or overvariability of the ensemble can be detected from the shape of the rank histogram.

Other evaluative techniques (spread/skill relationships, cluster analysis, etc.) may prove useful depending on the research issue in question.

Since the parameters verified should be driven by the needs of the end users of the forecasts, the group suggested that ongoing verification efforts begin to place more emphasis on sensible weather instead of traditional fields such as the 500-hPa height field. There should also be a concerted effort to assess performance of ensemble forecasts for rare events and for high-impact weather (e.g., severe thunderstorms). Of course, verification assumes that accurate observed fields are available, which unfortunately is not always the case for many regions (e.g., over the ocean), weather features (e.g., clear-air turbulence), and scales (e.g., meso- and microscale). In fact, the question of how to verify mesoscale EFs and validate model variability, in the absence of complete observations and quantitative estimates of observational/analysis uncertainty, is one of the most challenging issues facing the community.

Allocation of computational resources is another important consideration for optimal implementation of any analysis-forecast system. Ongoing evaluation will always be required to determine the optimal trade-off between grid spacing and the number of ensemble members; this can be expected to change as computational resources increase. The evaluation should also include the best ways to deal with model errors (see section 4).

The research community and forecast users would benefit from access to both the full ensemble forecast fields and some quick, useful summary information of the EFs. The full forecast fields would be useful to leverage the local forecast offices and universities to accelerate the development and improvement of ensemble forecast systems. These users could explore in full the potential benefits of ensembles and tailor ensemble products to their particular forecast problem, and would be provided the necessary initial analyses and lateral boundary conditions to run finescale, limited-area mesoscale ensembles on local workstations. Then again, in view of operational time constraints, it is impossible for a forecaster to examine and mentally synthesize every individual outcome from an ensemble, so synthesized products are desired in addition to raw model output. Further research into effective ways to condense, synthesize, and visualize ensemble output is suggested.

Closely related to the synthesis of ensemble output is its calibration. It is firmly established that statistical postprocessing of NWP output can significantly improve the skill of deterministic forecasts, primarily through the reduction of biases (e.g., Carter et al. 1989). The calibration of ensemble forecasts presents greater challenges than that of deterministic forecasts because the higher moments of the probability distribution may be mis-specified, not just the mean. Recent results indicate that techniques besides multiple linear regression can be successfully employed to calibrate ensemble output (e.g., Zhu et al. 1996; Hamill and Colucci 1998; Eckel and Walters 1998).

Discussion group participants included D. Baumhefner, T. Eckel, K. Mylne, M. Rennick, D. Richardson, S. Tracton, C. Ziehmann, and S. Mullen (chair).

6. Conclusions

Clearly, research on EFs and their use is growing. Despite the progress, ensemble forecasting is not yet used to its full potential in this country. We believe that this is due partly to a long institutional and societal inertia toward making and using deterministic rather than probabilistic forecasts, and partly because ensemble forecasting is still a relatively new endeavor. This workshop highlighted how far we have come in the last five years. Five years ago the potential usefulness of ensembles in data assimilation was not widely appreciated. Five years ago we knew model errors were problematic but had few ideas about how best to address them in an ensemble. Five years ago our knowledge of how to interpret ensemble forecasts and how to use them was minimal. Over the next five years, our community will be testing coupled ensemble forecast/data assimilation schemes in realistic numerical models; we will be researching and testing ways of addressing model errors and model uncertainty; and we will be looking for more effective ways to verify and communicate ensemble forecasts.

References


ABSTRACT

Among the many natural disasters that disrupt human and industrial activity in the United States each year, including tornadoes, hurricanes, extreme temperatures, and lightning, floods are among the most devastating and rank second in the loss of life. Indeed, the societal impact of floods has increased during the past few years and shows no sign of abating. Although the scientific questions associated with flooding and its accurate prediction are many and complex, an unprecedented opportunity now exists—in light of new observational and computing systems and infrastructures, a much improved understanding of small-scale meteorological and hydrological processes, and the availability of sophisticated numerical models and data assimilation systems—to attack the flood forecasting problem in a comprehensive manner that will yield significant new scientific insights and corresponding practical benefits.

The authors present herein a set of recommendations for advancing our understanding of floods via the creation of natural laboratories situated in a variety of local meteorological and hydrological settings. Emphasis is given to floods caused by convection and cold season events, fronts and extratropical cyclones, orographic forcing, and hurricanes and tropical cyclones following landfall. Although the particular research strategies applied within each laboratory setting will necessarily vary, all will share the following principal elements: (a) exploitation of those couplings important to flooding that exist between meteorological and hydrological processes and models; (b) innovative use of operational radars, research radars, satellites, and rain gauges to provide detailed spatial characterizations of precipitation fields and rates, along with the use of this information in hydrological models and for improving and validating microphysical algorithms in meteorological models; (c) comparisons of quantitative precipitation estimation algorithms from both research (especially multiparameter) and operational radars against gauge data as well as output produced by meso- and storm-scale models; (d) use of data from dense, temporary river gauge networks to trace the fate of rain from its starting location in small basins to the entire stream and river network; and (e) sensitivity testing in the design and implementation of separate as well as coupled meteorological and hydrologic models, the latter designed to better represent those nonlinear feedbacks between the atmosphere and land that are known to play an important role in runoff prediction.

Vital to this effort will be the creation of effective and sustained linkages between the historically separate though scientifically related disciplines of meteorology and hydrology, as well as their observational infrastructures and research methodologies.
1. Introduction

During the past 30 years, floods have—with virtue of their often rapid onset, occurrence in highly populated areas, and difficulty of prediction and warning—claimed on average 139 lives per year in the United States (NOAA 1994); this figure is now approaching 200 (NRC 1996). The property damage during this same period averaged some $1 billion per year (now approaching $2 billion), which is nearly twice that associated with tornadoes and hurricanes (NRC 1996). Approximately 75% of all presidentially declared natural disasters involve floods (Chapman 1992; Fread 1995; NRC 1996).

Numerous scientific and technological developments during the past several years have positioned the research community to make significant advances not only in its understanding and quantitative forecasting of precipitation, but also in determining the fate of precipitation upon its entry into the hydrological system. They include, but are not limited to, emplacement of the WSR-88D radar network; use of polarization diversity radars to discriminate hydrometeor type and location, including the development of automated algorithms and the deployment of a dual-polarization WSR-88D test bed; creation of algorithms for computing precipitation intensity from satellite microwave reflectance data; development and operational testing of storm-resolving models that use explicit microphysical parameterizations and are initialized with finescale radar and other observations; application of advanced statistical downscaling techniques to disaggregate radar-observed and modeled mesoscale precipitation to the fine resolutions needed by hydrologic applications; development of distributed hydrological and coupled meteorological–hydrologic models based on advanced physics as well as high resolution soil and terrain data; and the deployment of in situ soil monitoring sites at high spatial density as a means for comprehensive water budget accounting and the provision of soil moisture information to both atmospheric and hydrological models.

These and other advances in meteorology and hydrology now make possible the use of modeled and/or observed, spatially distributed and temporarily evolving precipitation fields in hydrological models designed to predict river discharge associated with heavy precipitation events. To take full advantage of these capabilities and, in particular, to develop new ones, it is important to recognize that the atmospheric and hydrological systems are intimately linked, and that any effort to create a research plan related to flood prediction must consider this coupling.

The U.S. Weather Research Program (USWRP) Prospectus Development Team on Hydrological Aspects of Weather Prediction and Flood Warnings (PDT 9) met in Honolulu, Hawaii, from 30 January to 1 February 1998 to discuss those areas in which coupled efforts in meteorology and hydrology could be expected to lead to major advances in one or both fields as related to heavy precipitation and flood prediction. The principal focus was high-impact meteorological events that have significant implications for the hydrological system, and for which forecasts plausibly could be improved through application of available technology in a coupled atmospheric–hydrological framework. These events typically involve severe flooding and include convective storms, snowmelt events, rain-on-snow events, orographically forced precipitation systems, and peak flow enhancement at specific localities due to frontal-induced phenomena and tropical storms and hurricanes following landfall.

The general charge to PDT 9 was the following.

- To identify and delineate critical scientific and technological issues related to linkages among quantitative precipitation estimation (QPE), quantitative precipitation forecasting (QPF), catchment runoff, and other land surface processes in a coupled meteorological/hydrological framework. Particular emphasis was given in this context to the prediction of floods and potential flood forecast situations on timescales from approximately 1 h to 10 days.
- To work toward establishing workshops or other community-wide activities that are deemed necessary for achieving the above goal, particularly with regard to research strategies, institutional linkages, and resource management.
- To draft a research prospectus that can be subjected to community scrutiny and published and disseminated to appropriate groups such as the USWRP Science Steering Committee and the National Academy of Sciences. This prospectus will serve as the blueprint for a hydrometeorological research plan related to flood prediction, and will be highly synergistic with its previous PDTs.

To deal most effectively with the broad spectrum of issues arising from this charge, the presentations and discussions were organized around the following principal themes:
2. Historical perspectives and current state of knowledge

a. Hydrology and meteorology in perspective

Members of PDT 9 entered the Honolulu meeting with a general recognition that, in the educational, research, and operational frameworks, meteorology and hydrology have evolved along independent and even divergent paths. Indeed, although this “gap” has been noted in numerous reports (e.g., NRC 1996), its depth and significance—especially in advancing the science and technology of flood forecasting—became increasingly evident throughout the meeting to the point where the research methodologies, models, and use of observations by meteorologists and hydrologists were sufficiently different as to warrant a discussion in this report. Thus, we devote this section to educating the reader about the historical as well as contemporary differences between meteorology and hydrology, and seek to identify key issues that must be addressed in order to evolve a coherent scientific attack on the flood forecasting problem.

The discipline of hydrology traces its roots to civil engineering, where topics such as river hydraulic structure design, flood plane zoning, reservoir operation, water availability, and ground water use and remediation received attention during the second half of the twentieth century. In the context of flood forecasting, hydrologists traditionally have relied upon simple conceptual models (e.g., unit hydrograph and linear reservoir theory) or simple physically based numerical models of both the lumped and distributed type (e.g., Freeze and Harlan 1969). These models have been justified in part by the fact that, unlike weather forecasting, in which both observing systems and atmospheric properties are distributed rather uniformly in time and space, hydrology suffers not only from a paucity of in situ measurements, but also from the need to characterize a geometrically complex physical system whose properties can change by several orders of magnitude over distances of only a few meters. To meet these challenges, hydrologists have relied heavily upon statistical techniques for site-specific model calibration, which is in sharp contrast to meteorologists, who traditionally have taken a largely deterministic, site-independent, and physically based approach to modeling.

Many hydrologists share the opinion that hydrologic models will continue to require calibration for some time because many properties of the catchment system are not yet directly observable. For example, hydraulic properties of the soil, the time-dependent role of vegetation, and the non-stationary behavior of channel characteristics (e.g., cross-section geometry and bed roughness, especially in the presence of extreme floods) have significant effects on predictions yet are difficult to measure at the appropriate scale (e.g., Beven and Sorooshian 1997). Indeed, calibration, and its dependence upon scale, are important stumbling blocks to successful hydrologic modeling and
flood prediction that the recommendations proposed herein address. Additionally, the search for scaling properties of hydrologic variables, their relationship to the scale of measurement as determined by the available technologies, and their effect on a variety of applications are important challenges for both research and operational hydrology.

During the past two decades, meteorology and hydrology have increasingly faced common challenges, driven in large part by scientific necessity. One of the most important examples involves the origin and fate of precipitation. Although precipitation is the main forcing variable for surface hydrologic processes, it is poorly predicted by meteorological models, even at space and time scales much coarser that those necessary for flood and flash-flood forecasting. New efforts are under way, as described below, to improve quantitative precipitation forecasting in models and to utilize this information effectively in hydrological models.

Significant advances have been made recently in the spatially distributed numerical models of rainfall-runoff processes. These models take into account detailed variations in topography, in hydraulic properties of the soil, and even in vegetation that exists within a basin and governs the surface and subsurface water flow. They also are capable of describing the time and space evolution of soil moisture, which now is recognized as vitally important to meteorological predictions. On the other hand, the main input to such models is precipitation, which traditionally has been estimated from rain gauges. In the United States, the national network of WSR-88D radars, with the capability of providing high-resolution estimates of accumulated precipitation, has the potential to play a major role in driving future developments and operational applications of distributed hydrological models in much the same manner as it is now doing for mesoscale numerical weather prediction models (Droegemeier 1997).

Meteorologists now recognize more than ever before the complex issues associated with atmospheric variability and uncertainty and are applying more sophisticated statistical techniques to both numerical modeling (e.g., ensemble forecasting, spatial downscaling) and the observations that define the model initial state (e.g., radar-based estimates of precipitation rate). Both meteorologists and hydrologists have developed techniques for assimilating observations (e.g., McLaughlin 1995; Ghil et al. 1997), including data from radars and other remote sensing systems. At the present time, however, few such methods have been included in operational hydrological models. Additionally, although several efforts are under way to couple atmospheric and hydrologic models (e.g., Georgakakos 1986; Shuttleworth 1996), this area of research remains relatively new, as does the application of data assimilation techniques to coupled models.

One interesting difference between meteorology and hydrology is that the former has a long and successful history of evolving major numerical models for use by the entire community (e.g., MM5, RAMS, ARPS, and the climate models supported by NCAR), while many of the hydrological models are site-specific and thus used in a more limited manner. Indeed, the mesoscale meteorology codes are serving as the foundation for a new community-wide dual research and operational forecast system known as Weather Research and Forecasting (WRF; Dudhia et al. 1998). Somewhat remarkably, more than 25 groups in the United States are applying mesoscale atmospheric models with varying degrees of sophistication in daily real-time forecasts. Recently, the hydrological community also has begun moving toward a more unified approach to modeling and testing, partly because of the need to understand the global hydrologic cycle, and partly because the optimal use of limited water resources necessitates a greater use of remotely sensed data and the development of hydrologic models for data-sparse regions.

The hydrological community is now assessing strategies to effectively deal with these new challenges (Entekhabi et al. 1999). The ready availability of remotely sensed data, coupled with joint hydrological and meteorological field programs and improved representations of land–atmosphere coupling in atmospheric models, have set the stage for significant advancement in numerical precipitation and flood forecasting. An implicit goal of the recommendations provided herein is to support a community focus in hydrological and hydrometeorological research and modeling toward the use of physically based distributed models, along with the integration of deterministic modeling and statistically based approaches.

b. Hydrology overview—Observations and processes

The principal observing tools used by hydrologists to characterize the movement of water over a catchment pertain to the main input and output. The input is rainfall, which traditionally has been observed as point values provided by rain gauges. These data are interpolated to form estimates of mean areal precipitation for the entire basin or its elements (subbasins
or grid boxes). Recently, weather radars have begun to provide quantitative estimates of accumulated precipitation. Although the value of radar data is indisputable (e.g., Obled et al. 1994), particularly with regard to space and time resolution for hydrological models, the quantitative characteristics of its uncertainty sources are not well known but must be established for use in both meteorological and hydrological models.

The main output of interest to hydrologists is discharge at the outlet of a basin, which typically is measured by stream or water stage gauges. The former provides information on volumetric flow rate, the representativeness of which for a point measurement in a channel cross section must be taken into account, while the latter records water elevation. Both are important for determining stream discharge rate, which is the key quantity used in flood forecasting. Surface runoff throughout the basin, and the discharge at outlets of subbasins, typically are not observed.

Equally important to hydrology are the water content and hydraulic properties of the soil, along with the vegetation state. Initial soil moisture, the infiltration capacity of the soil at the surface, and the relative position of the ground water table determine surface runoff, two modes of which are widely recognized. In the first, surface runoff occurs when the rain rate exceeds the soil infiltration capacity. In the second, the soil column becomes saturated when the groundwater table rises all the way to the surface. Both modes may be present in a single basin, although typically certain soil types and hydrogeologic and climatic conditions favor one over the other.

Most hydrologic processes are tightly coupled to a local landscape that can vary significantly at small scales. Consequently, for surface runoff, the impact of small-scale surface structures tends to be significant because the local flow is directed by it as the water moves down slope. If the rainfall rate were uniform over hundreds of kilometers, the effects on hydrographs (time series of stage or discharge rates at a point in a stream or river) of the small-scale landscape structures could probably be treated via similarity and scaling theories (e.g., Sivapalan et al. 1987). Unfortunately, this is rarely the case. In fact, at the 1–10-km scale, spatial and temporal structures of the precipitation field, and the topographic structure of the networked landscape, begin to interact in a manner that exerts a profound influence on the shapes of the hydrographs downstream. Needless to say, a similarly important effect is played by the subsurface “landscape,” but it is even more difficult to measure directly. Thus, if tributary streams and rivers are phased such that their discharge is additive under conditions of heavy precipitation, with the phasing a function of the scales and motions of structures in the precipitation field as they move across the terrain, then a severe flood becomes possible.

c. Hydrology overview—Numerical models

Hydrologic models can generally be classified as distributed or lumped. The former use spatially distributed grids, usually situated over one or more basins, to represent, with varying degrees of sophistication, the spatial and temporal variations in the physical system both above and below the ground. They use as input precipitation estimates from radars, atmospheric models, or combinations thereof and output estimates of runoff, infiltration, and evapotranspiration. Lumped models take a more limited view of the basin by simplifying the detailed hydraulic characterizations into representations of bulk properties on the large scale of regions containing considerable heterogeneity. They seek parametric relationships between the given rainfall input and the discharge output that minimize some measure of fit to discharge observations.

Once calibrated, existing flood forecasting models must be used with data of the type and scale with which they were calibrated. Inputting accurate information on smaller scales, as is the case with WSR-88D radars replacing rain gauge based estimates of input (e.g., Foufoula-Georgiou and Krajewski 1995), can degrade the output and lower the skill, even though the small-scale information might provide a better characterization of the actual physical system. Consequently, lumped models should be completely recalibrated every time the scale of the input data is changed. Applying this idea to the entire country is a monumental task. Clearly, innovative approaches to the problem are called for.

Prior to the advent of comprehensive radar coverage and mesoscale models with grid spacings of a few kilometers, there existed little incentive to develop and apply detailed, process-based flood forecasting models. This led to highly parameterized models, calibrated using historical rainfall and discharge records, which ignored both the terrain characteristics and the spatial variability of the rainfall. Distributed models, on the other hand, are designed to take into account the spatial variability of topography and all other relevant inputs and parameters. Therefore, such physically based distributed models, in principle, require no
calibration because they attempt to predict physical processes based on observable parameters at scales on which the processes operate. Unfortunately, many crucial parameters vary at scales much smaller than those resolvable with current observing technology. Observations of inputs are subject to significant uncertainty which, when combined with the highly nonlinear structure of hydrological processes, leads to uncertainty in the outputs.

Finally, hydrological models presently are handicapped by the limited understanding of certain aspects of water flow throughout a basin. This situation is exacerbated by the fact that model forecasts are verified in a very limited manner, that is, by comparing discharge rates at the outlet of a basin input. Innovative approaches to distributed model validation are required in order to reduce the ability of "tuned" models to obtain the correct solution for the incorrect physical reason. Future improvements in remote sensing technology, especially in the context of soil moisture (e.g., Basara et al. 1998), channel characteristics, and flow, coupled with more sophisticated geographic information systems and computing capabilities, will undoubtedly translate into improved predictive capabilities of distributed hydrological models.

d. Meteorology overview—Radar observations

Never before has the opportunity to provide effective meteorological input to distributed hydrologic models, and especially to flood prediction models, been so ripe with scientifically interesting possibilities and achievable goals. One key to attaining these goals is the availability of nearly contiguous single-Doppler radar data from the WSR-88D (NEXRAD) network. However, even though this network has vastly improved the detection and short-term warning of severe weather, considerable work remains before it can produce estimates of surface rainfall with the resolution and accuracy required by distributed hydrological models. Some of the important issues follow.

Precipitation estimates from the WSR-88D are affected by many factors including the calibration of individual radars, attenuation (although this is minimal for S-band radars, like the WSR-88D, in most situations), ground clutter, beam blockage, beam broadening with range, siting (especially at high altitudes in significant terrain, which can lead to underestimation of precipitation, especially for shallow systems), and brightband contamination. Perhaps the most fundamental contributor to uncertainty is the lack of a unique Z–R (reflectivity–precipitation rate) relationship that relates backscattered power to instantaneous rain rate. The Z–R-based estimates are especially problematic in climatological regimes with highly variable rainfall characteristics (e.g., over the eastern slopes of the Rocky Mountains, which can exhibit high-based continental convection one day and moist, monsoonlike precipitation the next).

The contamination of WSR-88D reflectivity by hail and other frozen species, especially over the Great Plains, also is problematic. The radar processing software deals with these situations by truncating reflectivity, typically at a value of 53 dBZ. Under certain conditions, a severe underestimation of actual rain rates may result since reflectivities associated with rain-only situations can exceed the threshold value.

The systematic calibration of the WSR-88D, coupled with research to refine the choice of Z–R relationships to account more accurately for variability in the rainfall regime (which is fundamentally linked to variations in drop-size distributions), will permit full realization of the rainfall-mapping capabilities of the radar. Concurrently, research continues to improve the estimation of both rain and snow precipitation rates and to identify precipitation regimes (rain vs hail, rain vs snow) using multiparameter (dual-polarized) radars (e.g., Vivekanandan et al. 1999). The dual-polarized rainfall estimators (especially those using specific differential phase) are markedly less sensitive to variations in drop-size distribution produced by changes in the rainfall regime at a given location than are their Z–R (WSR-88D) counterparts. Furthermore, the estimators are less prone, and in some cases are even immune, to power calibrations, partial beam filling conditions, clutter, and the presence of frozen hydrometeors.

Research is presently under way to improve the estimation of WSR-88D-based precipitation rates and hydrometeor identification via the collocation of operational and dual-polarized research radars [e.g., CSU’s CHILL, NCAR’s S-pol, and the National Severe Storms Laboratory (NSSL) Cimarron]. Efforts also are under way at National Oceanic and Atmospheric Administration (NOAA)/NSSL to identify optimal strategies for upgrading existing WSR-88Ds to dual polarization capability. Estimation of rainfall rates would be improved significantly by the availability of dual-polarization capabilities on the WSR-88D radars in that region.

All radars are limited in regions of significant terrain due to beam blockage. Vertically pointing radars, however, show significant promise in high terrain for
estimating rainfall rates and identifying rainfall regime (e.g., convective vs stratiform). Future field work aimed at merging radar estimates of precipitation rates with hydrologic models should include this technology. Short wavelength dual-polarized radars (e.g., X-pol, and the X-band radar operated by NOAA/ETL) are well-established technologies that extend the dual-polarization estimation of precipitation to low rain rates by virtue of larger propagation differential phase shifts at X-band versus S-band (a factor of 3 for a given rain median). Short wavelength radars also afford a smaller sample volume and are therefore suitable for application to small-scale structures in the precipitation field. However, they are correspondingly more vulnerable to attenuation than S-band radars.

e. Meteorology overview—Numerical models

Research meteorological models are now being tested by the academic and other elements of the research community at resolutions of a few kilometers, and operational centers are using such resolutions to support special operations (e.g., the Olympics, military campaigns). Indeed, more than 25 groups in the United States alone are operating such models in real-time, in most cases via collaborations with local National Weather Service Offices or NOAA Cooperative Institutes. Never before has the general research community, especially within academia, taken such an active interest and played such an important role in operational meteorology. Such efforts have undeniably been spurred in large part by the ready availability and low cost of significant computing power, gridded data, and display resources. Although many of these projects are not formally funded, they are serving as the basis for valuable research in predictability, data assimilation, ensemble forecasting, and numerical technique and physical parameterization development and testing.

Perhaps the greatest challenge for numerical models, and certainly one of the most important in the context of flood prediction, is quantitative precipitation forecasting (QPF). As noted by Fritsch et al. (1998), QPF skill by operational models remains relatively low, particularly in light of recent vast improvements in observational and modeling technology, and especially for extreme events. Not surprisingly, the skill levels are lowest during the warm season because over 80% of significant warm season precipitation is associated with thunderstorms (Heideman and Fritsch 1988), for which operational models tend to perform poorly.

At the grid spacings now being tested by the research community, model precipitation fields are approaching the scales of practical relevance to hydrologic models. However, such resolution enhancements bring new challenges in the fundamental application of numerical models to the atmosphere. First, in situations where topographic and other forcings are relatively weak, high spatial resolution in the model will likely require observations of similar resolution. For example, Droegemeier et al. (1999) showed that, without the retrieval and assimilation of WSR-88D base data, an observed supercell hailstorm was essentially absent 2.5 h into a forecast using 3-km resolution. The use of WSR-88D radar data to initialize storm-resolving atmospheric models was the scientific underpinning of the Center for Analysis and Prediction of Storms (e.g., Droegemeier 1997) and has now become a research focus of other efforts, particularly the WRF model development project (Dudhia et al. 1998). While the lack of real-time base data from the WSR-88D network has inhibited such projects, efforts are now underway to acquire and use such data from numerous radars in explicit real-time storm prediction testing (Droegemeier et al. 1999).

A second consideration is that high model spatial resolution provides for the explicit representation of clouds and their attendant microphysical processes (as opposed to implicit treatments via convective schemes), but also requires appropriate observational data for both initialization and verification. While many sophisticated microphysics schemes exist and are being used successfully in research applications, operational limitations usually require much simpler forms. The simplification process is not well posed, and only to a limited extent has it been undertaken in consideration of the potential for using model precipitation fields in hydrologic models. To complicate matters, no generally accepted methods have been put forth for verifying model-generated precipitation forecasts of highly intermittent events (e.g., individual convective storms), or for establishing the physical realism of the predicted precipitation structures compared to radar data, for example. These areas are central to both meteorology and hydrology and are included in the recommendations herein.

Finally, at high spatial resolutions, horizontal variations in vegetation type and state, soil type, and soil moisture can have a profound influence on the initiation and subsequent evolution of intense precipitation-producing weather events. Once again, the inclusion of advanced representations for these properties and processes in atmospheric models must be predicated on a solid understanding of the underlying
physical behavior, particularly exchange processes between the atmosphere and ground surface, and be accompanied by suitable input as well as verification data between the atmosphere and ground surface, and be accompanied by suitable input as well as verification data (e.g., Chen et al. 1997). Fortunately, new satellite instruments are making possible a much more complete and quantifiable analysis of the ground surface state, and in situ measurements of surface heat and moisture fluxes from large mesonetworks and instrumented towers (e.g., Brotzge et al. 1999) is advancing the development, calibration, and verification of suitable surface physics parameterizations. A note of caution is in order, however. With increased detail in any physical or dynamical process comes the possibility of decreased predictability owing to the greater uncertainty in observations and weaker fundamental understanding of small-scale processes (e.g., Lorenz 1969).

3. Classification of floods and related scientific challenges

a. General comments and prediction scenarios

Floods are notably diverse from both meteorological and hydrological points of view. Some result from prolonged heavy precipitation over an extended region while others are shorter in duration and more local in origin. The focus of this report is on floods produced by heavy precipitation events for which better atmospheric predictions are expected to yield improved hydrological forecasts. For seasonal floods, which might be considered more a part of the climate domain, snowmelt runoff or persistent rainfall can set the stage for flooding when shorter-term, intense precipitation occurs over wet terrain crossed by rivers and stream networks with bankfull or overbank flows.

In the western United States, floods of record often occur in early summer when the seasonal snowmelt peak is delayed by cold, cloudy spring weather followed by an unusually early period of intense summer convective rain. Prolonged seasonal rainfall can establish antecedent hydrological conditions for which additional precipitation can produce severe flooding. In such cases, accurate local flood prediction depends upon both an accurate precipitation forecast and an accurate hydrological characterization of the antecedent conditions. In fall or late winter, the antecedent precipitation may have fallen as snow, and the antecedent hydrologic condition may consist of a layer of saturated snow covering a broad region of frozen or saturated soil that is susceptible to impact by intense warm rain. This scenario is common in the Coast and Cascade ranges of Washington and Oregon, and in the Sierra Nevada of California.

From a prediction point of view, the flood problem as treated in this report consists of five primary components: 1) a precipitation prediction, 2) an antecedent soil moisture prediction, 3) a runoff calculation, 4) a flood routing calculation, and 5) a river stage prediction. The latter four are often coupled in a single hydrologic model, but for the purposes of the present discussion are best thought of as independent entities.

Precipitation can be predicted by a high-resolution meteorological model or inferred from WSR-88D or other radar data. In the former case, knowledge of the three-dimensional water vapor distribution is of paramount importance, and satellites, commercial aircraft, and ground-based Global Positioning System (GPS) sensing systems hold promise for improving both the spatial and temporal resolution of water vapor information. Soil moisture and related quantities must be obtained from a surface moisture (accounting) hydrologic model, or from a soil model embedded within an atmospheric model. These surface and subsurface models typically operate on timescales longer than those of atmospheric models, but also on a finer spatial scale. They require proper treatments of plant physiological processes and evaporation, both of which can vary considerably with the local topography.

Runoff calculations are the domain of hydrologic models and require input from a surface moisture model; however, runoff models need to operate on a much shorter timescale for many if not most flood problems. Coupling the moisture accounting and runoff calculations has traditionally been detrimental to both because of the substantial process and timescale differences. Flood routing models for gauged basins could be improved by the use of better predictive algorithms of hydraulic properties in stream channels, and by more highly resolved terrain and vegetative features.

b. Heavy warm season convective rainfall

Floods associated with convective rainfall in flat and hilly terrain are characteristic of the midwestern United States. In such conditions, floods on large rivers are associated with spring melting, combined with considerable deepening of the high springtime flows in response to upstream convective precipitation. For rivers and streams in smaller basins, flooding is associated with heavy or sustained precipitation in all or part of the basin. An example of the latter occurs in the upper Walnut River, situated in rolling terrain just
east of Wichita, Kansas. The Walnut River occupies a basin of about 60 km across and 100 km in length. At Augusta, which is situated at the center of the basin, the Walnut River has reached flood stage during about half of the last 40 years. The watershed drains at Winfield, where the Walnut has reached flood stage in about one-third of the last 100 years, with more frequent flooding in recent years. At both locations, most of the flooding is associated with convective precipitation in the spring and summer. For example, in the Washita River, Oklahoma basin, singular spectrum analysis in progress suggests that land use changes (irrigation, grazing, reservoirs, pumping, and baseflow) are a major factor in the interannual to decadal and longer changes in runoff patterns (unpublished research). The relative effects of anthropogenic change from terrestrial and atmospheric sources provides a significant challenge to numerical models.

Local soil conditions and synoptic forcing both play a role in determining the location and intensity of precipitation. Surface fluxes influence the structure of the boundary layer and determine when and where convective temperature will be reached. Differential heating and evaporation associated with adjoining dry and moist soil conditions can enhance the likelihood of precipitation either locally or downwind. Finally, preexisting wet soils can increase both the likelihood of precipitation (by increasing evaporation from the soil and evapotranspiration from the plant cover) and the likelihood of flooding once precipitation occurs. For accurate precipitation and flood prediction in this situation, mesoscale models need to include a reasonably comprehensive representation of current soil and vegetation conditions. Similarly, local forecast centers will need to track soil moisture conditions if flash floods are to be predicted accurately from WSR-88D output.

Although the discussion so far has emphasized intense, short-term precipitation events, it is important to not overlook situations where persistent precipitation occurs over a large area (e.g., as in the recent El Niño in southern California). In such cases, flooding results not from any particular storm, but rather from the collective effects of a succession of storms.

c. Intense convective precipitation in the Front Range

During a typical summer in Colorado, intense convective storms propagate from the mountains toward the plains. The Fort Collins, Colorado, flash flood of 1997 (Petersen et al. 1999) was an outcome of one such storm. Under conditions of weak vertically averaged winds over the mountains, these storms can stall, thereby delivering their precipitation to a small or moderate sized drainage basin and resulting in severe flooding. In these cases, the top soil is usually dry and the low-order tributaries also are dry or nearly so.

d. Rain on saturated snow

Prolonged periods of heavy rain on saturated snow frequently cause severe flooding in the Pacific Northwest and northern California. Warm fronts laden with moisture move onto the coast from the Pacific Ocean during the late fall and winter seasons, and it is not uncommon for these events to follow periods of heavy wet snowfall. Extratropical cyclones also can cause heavy precipitation and flooding, for example, as a consequence of El Niño. In particular, moisture-laden air transported from the eastern Pacific Ocean can lead to very intense rainfall in places usually devoid of this type of flooding.

e. Tropical cyclones

Although the mechanics of heavy precipitation associated with tropical cyclones do not differ significantly from those associated with moisture-laden extratropical cyclones, the amount of water available for delivery in a very short period of time is much larger, and the precipitation often falls on low coastal regions that are poorly protected from severe flooding, that are overbuilt, and that often are difficult to evacuate. Typically, such flooding also results in a substantial degradation of local water quality and in some areas may have an adverse effect on the coastal fauna and flora. The rain is of a warm tropical type and precipitation fields probably can be resolved with reasonable accuracy using the WSR-88D network. Because the disturbances move quite rapidly, even after landfall, and contain exceptionally heavy precipitation, monitoring the heavy rainfall and associated flooding requires considerable use of existing instrumentation. These platforms could be supplemented by highly portable and rapidly deployable devices (e.g., the Doppler on Wheels; Wurman et al. 1997).

4. Recommendations

The previous sections have introduced the challenges facing scientists in their quest for improved flood predictions. With that in mind, PDT 9 provides the following specific recommendations. We divide them broadly into three categories that emphasize
(a) improvements in the quantitative estimation of precipitation; (b) improvements in numerical models and their associated elements; and (c) the creation of “natural laboratories,” each having specific meteorological and hydrological characteristics, in which to undertake intensive field experiments over sufficiently long periods of time.

a. Quantitative precipitation estimation

Recommendation 1: Improve algorithms for the radar-based estimation of quantitative precipitation and establish measures to quantify uncertainty.

An accurate quantitative estimate of surface precipitation (QPE) is one of the most important elements in hydrologic modeling and flood forecasting. As noted by PDT 8 (Fritsch et al. 1998), improved algorithms are needed to estimate precipitation from the WSR-88D radar over diverse climatic regions. Efforts to discriminate precipitation type from dual-polarization radars also must be enhanced. In addition to quantifying the spatial, temporal, and amplitude accuracies of precipitation estimates using all available observations, efforts must be directed toward establishing the associated statistical uncertainty (e.g., representativeness of point measurements). It makes little sense to consider forecast uncertainty until observational uncertainty has been determined. Ultimately, probabilistic estimates of QPE likely will be required, and indeed represent the most natural framework for this particular quantity.

Recommendation 2: Develop techniques for blending data from multiple sensors.

The most reliable QPE will be obtained by combining, in an objective fashion, observations from multiple sensors. The current National Centers for Environmental Prediction “Stage IV” precipitation analysis, which calibrates WSR-88D radar estimates of precipitation with surface gauge reports, is an important first step in that direction. The addition of satellite and other observations, along with improved algorithms for combining multisensor output, perhaps of the variational type, is viewed as critical to the ultimate success of QPE.

The use of rain gauge data in evaluating the accuracy of radar- as well as mesoscale model-based precipitation forecasts has been hampered by the fundamental sampling-scale differences between these systems. Considering that the surface area of a typical radar pixel and a rain gauge orifice differ by about 8 orders of magnitude, efforts are clearly needed to quantify the natural variability of rainfall between them. This will enable estimation and prediction of errors, and will require that clusters of rain gauges having separation distances ranging from 10 to 1000 m be placed at strategic locations in the radar coverage pattern.

Recommendation 3: Establish a community infrastructure for collecting and making available large datasets of remote and in situ hydrometeorological observations and products.

Both meteorological and hydrologic research in general, and flood research in particular, are severely hampered by the lack of inexpensive and easily obtainable high-resolution rainfall and other observations during periods of intense precipitation. For example, the WSR-88D radar network provides comprehensive estimates of rainfall-rate fields at high spatial resolutions so that, in principle, these data could be used in combination with digital elevation maps over large and hydrologically important areas of the United States for flood prediction and water management. Unfortunately, NOAA data management procedures have created a cost barrier that prevents the use of these data for most hydrologic purposes. A mechanism needs to be created to make valuable radar information available to researchers at reasonable cost and time delays, and initial efforts by the research community to do so (Droegemeier et al. 1999) represent a positive first step and should be continued.

In order for the hydrological and meteorological communities to effectively attain the goals outlined herein, efforts should be directed toward building a library of long and complete datasets in several judiciously selected locations. The need for long duration is driven by the response time for the soil and basins, as well as the need to capture a wide variety of weather system types. These datasets should be composed of both remotely sensed data, for example, from radars and satellites, as well as measurements from surface sites and instrumented towers. The considerable spatial variability in rainfall, coupled with the non-Gaussian nature of the errors, is a strong argument for the availability of long time records. Finally, owing to the intermittent nature of rainfall and its paucity of in situ gauge observations, long-term records (i.e., many years) are needed to produce sufficiently large samples, for statistically significant evaluation of model and algorithm improvements under a range of hydroclimatological conditions. In the context of coupled hydrometeorological model-based flood forecasting, event-based verifications are simply inadequate, especially considering the need for continuous monitoring of soil properties.
Recommendation 4: Enhance the current in situ hydrological observing networks in the United States. As noted by Entekhabe et al. (1999), the landsurface and subsurface observing infrastructure, like so many others, has been diminished over the past few decades owing to political and fiscal issues. Indeed, they point out that less is known today about many important parameters than in the recent past—a notable irony since new observing systems allow for the ready measurement of important quantities such as soil moisture, evapotranspiration, and surface fluxes. It is therefore vitally important that steps be taken to stabilize and improve the observational infrastructure. USWRP funds should be directed toward this specific need.

b. Numerical modeling

Recommendation 1: Investigate the requirements of and strategies for running atmospheric and hydrologic models in a coupled manner.

Many in the scientific community agree that coupled atmospheric–hydrologic models are the wave of the future. Indeed, such efforts have now begun (e.g., Bonan 1996). However, a number of important issues must be addressed before such efforts can be expected to succeed, including the degree of coupling required, and strategies for dealing with the different process timescales inherent in atmospheric and hydrologic systems. Additionally, atmospheric models currently generate output at resolutions much coarser than is nominally required by hydrologic models. Consequently, the use of statistical downscaling techniques should be investigated to see if they provide useful information at the required resolutions. Indeed, such studies may suggest a practical lower limit for atmospheric model resolution. Finally, to facilitate uncertainty estimation for flood forecasts, uncertainties in atmospheric model precipitation output must be complemented with a characterization of their errors.

Recommendation 2: Conduct sensitivity and parameter estimation studies of hydrologic and atmospheric models run individually and in a coupled manner.

Systematic studies are needed to determine which aspects of both hydrologic and atmospheric models (e.g., representation of land cover, methods of using input data), run individually or in a coupled fashion, exhibit the greatest sensitivity as a means for identifying those components and physical processes that should receive the most attention and presumably yield the greatest scientific payoff. Observing system simulation experiments and adjoint-based sensitivity approaches are ideally suited for this task, and for testing the impact of and strategies for using “observations” that are not yet available (e.g., high-resolution precipitation fields from deterministic atmospheric prediction models). The effect of scale parameter estimation, along with the sensitivity of hydrologic model output to spatial and temporal variations in precipitation input, are especially challenging and relevant problems, as is the sensitivity of atmospheric models to land–atmosphere exchange processes.

It is an axiom of computational fluid dynamics that one never has sufficient computing resources to achieve the spatial resolutions desired for a given problem. While the notion of “higher resolution will produce better results” is still generally true in most meteorological and hydrological applications, efforts must be directed toward investigating resolution sensitivity, particularly in coupled models. There remains considerable uncertainty regarding the detail actually needed, particularly in observations but also in model grids, in the context of specific problems. This is particularly true for hydrologic models, which have not enjoyed the long history of increasingly finer resolutions associated with atmospheric models. The implications of this issue are significant, and could lead to a more rapid implementation of models and methods in light of available computing resources. For example, if statistical downscaling can yield appropriate probabilistic estimates of surface precipitation from relatively coarse-resolution atmospheric model output, then the need to run atmospheric models at expensive high resolutions may be averted, at least in some cases. Relationships between model resolution and the physical scales being represented also need to be studied, particularly in hydrology. The resulting information also would be very useful in guiding the deployment of new sensing systems (Ciach and Krajewski 1999).

Recommendation 3: Conduct extensive verification studies of atmospheric and hydrologic models, with emphasis on using the latter to verify the former. Verification is a sine qua non of forecasting, and in contrast to operational atmospheric prediction models, no long-term systematic verification records exist for operational hydrologic models. Forecast quality is the relationship that links observations to forecasts, and relationships between quality and value are nontrivial and can be counterintuitive. Consequently, efforts must be directed toward verifying forecasts produced by hydrologic models across broad parameter spaces, and care should be taken to avoid boiling down the information to one or a few statistical mea-
sures. Rigorous verification of atmospheric precipitation forecasts made at high spatial resolution (down to 1–3 km) also must be a goal, particularly with regard to the uncertainty described in recommendation 1. Specific measures of skill, and a determination of the skill actually required for a coupled atmospheric–hydrologic modeling system, must be established and tested using both simulated and observed data.

Recommendation 4: Improve the assimilation of observations into hydrologic models.

Hydrologic models frequently are applied to a single event. To extend their use for continuous simulation, efforts must be directed toward a more effective use of observations and other inputs (e.g., from atmospheric models) via data assimilation strategies. Some of the greatest improvements in atmospheric prediction have come from data assimilation, as it offers numerous ways for bringing observational error statistics directly into the forecast system. In hydrology data assimilation, techniques based on Kalman filtering have been used successfully with lumped and semidistributed models. These techniques need to be extended to detailed distributed models for assimilation of remotely sensed and other data. Additionally, other advanced approaches (e.g., adjoint, 3D variational assimilation) provide a coupled modeling framework for using the quantities actually measured by observing systems (e.g., radar reflectivity).

Recommendation 5: Assess the suitability of current atmospheric model microphysical parameterizations for use in hydrologic forecasting.

Cloud microphysical parameterizations in atmospheric models run the gamut from very simple and computationally efficient to extremely elaborate and tremendously expensive. All methods, however, including both implicit and grid-scale explicit schemes suffer from two problems: empirical formulation and difficulty of initialization and verification using observed data. Because precipitation, along with solar radiation, surface properties, and wind, is one of the most important inputs to hydrologic models, we recommend that considerable emphasis be placed on methods for initializing and validating microphysical process schemes in atmospheric models. In particular, comparisons must be made between the spatial structures of both model-generated and observed precipitation fields, and verification of surface estimates—especially at high spatial resolution and including measures of uncertainty—should be rigorously pursued. The use of dual-polarization radar data to initialize and verify moisture and precipitation fields of atmospheric models also must be expanded.

Recommendation 6: Combine statistical and deterministic approaches in modeling atmospheric and hydrological processes.

When atmospheric models are unable to resolve the space–time structure of precipitation at scales of hydrologic importance to flood prediction, a combination of deterministic and stochastic approaches that statistically parameterize subgrid scale rainfall might be advantageous. For this reason, characterizing the multiscale space–time rainfall variability of observed precipitation and testing its ability to be reproduced by models at different scales is a high priority. This approach to model verification, which is based on assessing the model’s ability to capture important space–time dynamics of observed storms, is distinctly different from current approaches that use much simpler verification metrics such as threat scores and rms error. It is expected that multiscale space–time measures of forecast performance will provide more useful feedback for model improvement as well as provide the means for assessing precipitation predictions at scales different from the model grid size but relevant to the hydrologic applications.

Recommendation 7: Improve the characterization of surface and subsurface properties and physical processes in atmospheric and hydrologic models.

Land surface properties and associated physical processes are vitally important to both meteorological and hydrological models. In particular, soil type and moisture content, as well as vegetation cover and state, play very important roles in the prediction of heavy precipitation events, particularly those occurring at or being forced by small scales. With high spatial resolution and coupled models comes an increased need to provide accurate specification of the land surface features and energy exchanges, as well as soil properties not only at the surface but in the subsurface as well. In many cases, parameterizations developed for use in coarse-resolution models are unsuited for application at smaller scales. Consequently, efforts should be directed toward improving the treatment in both atmospheric and hydrologic models of surface and subsurface characterizations, and toward using new datasets (e.g., from satellites) in place of the notably limited data now available (e.g., NDVI) to accommodate multiple vegetation and soil types within a single grid zone. The importance of scales at which variations in surface and subsurface properties impact flooding must be explored, and linkages to the ecologi-
cal community should be established as part of this process.

c. Natural laboratories

In order to test a model of an important natural phenomenon, a site is required at which (a) that particular phenomenon frequently occurs as a dominant signal, and (b) a comprehensive set of instruments capable of resolving all important aspects of the phenomenon can be assembled for an appropriate period of time. Additionally, it is advantageous if such sites already are part of a field campaign, or are situated near universities that have programs in hydrology and/or meteorology. Such sites might appropriately be termed “natural laboratories.” Although many high quality atmospheric and hydrologic datasets now exist, their measurement, coverage, or duration are not sufficient to allow for comprehensive studies of the coupled atmospheric-hydrologic system. The natural laboratories proposed below seek to overcome from these limitations.

Recommendation 1: Utilize “natural laboratories” for studying a variety of natural phenomena in the meteorology–hydrology coupled system

For specific types of floods, there exist particular sites at which the aforementioned criteria can be satisfied. Moreover, the need to test and improve all components of flood prediction models is so great that we strongly recommend establishing a set of natural laboratories for these purposes. The phenomena of concern here are of a scale that exceeds the typical watersheds of the Agricultural Research Service and involve terrain and land surface features that have not been dealt with in most larger-scale watershed investigations. Consequently, new sites of appropriate size and variety are needed.

In the context of floods, a reasonably large area must be examined in order to insure a suitable occurrence interval within the domain. This makes instrumentation of the domain more difficult and requires using one or more radars to capture the spatial distribution of precipitation. In addition, a reasonably dense river gauge network would be required to verify the locally predicted hydrographs. It would most likely be composed of both permanent United States Geological Survey as well as temporary stage gauges. Land surface information (such as vegetation type and state) could be obtained from satellite images calibrated to measurements made at judiciously placed surface stations. Means of extracting soil moisture content from plant imagery, surface energy exchange, and plant physiological models are available but need to be developed further. Techniques that relate the spatial structure of land surface variables to characteristics of the terrain obtained from analyses of digital elevation maps also will need to be employed.

Recommendation 2: Hold a workshop to establish the logistics and scientific framework for the “natural laboratories.”

The launching of any natural laboratory initiative should begin with a workshop that takes the information provided below as a starting point and evolves a detailed scientific plan with testable hypotheses and defined time lines. We therefore present below a discussion of possible sites based upon the phenomena to be studied and the unique geophysical settings required. This list is by no means exhaustive, but should be viewed as the starting point for a national workshop.

1) Site 1: Floods caused by intense rainfall from topographically induced summer convection

For heavy summer convective rain at least two types of natural laboratories are necessary. The first would deal with the types of floods that originate in the Front Range of Colorado, where moisture-laden convective cells propagate slowly because of weak winds aloft and become topographically trapped in relatively small drainage basins. Owing to the large number of such basins, flash floods of this type are sufficiently common to make waiting for one a viable option. Furthermore, the general meteorological conditions that foster this type of heavy precipitation can be identified easily, triggering a more intensive watch for particularly likely storms. When not on duty elsewhere, the CHILL and S-pol radars would presumably be available to observe a great expanse of the Colorado Front Range from their locations on the plains.

Because most of the Front Range streams are used for irrigation or to supplement municipal water supplies, a reasonably dense net of stream gauges exists and could be used for testing flood predictions. For example, more than seven stage gauges are positioned in the Boulder Creek watershed above Boulder for water management and flood warning. These Front Range gauges also provide sufficient information on the preflow flow in the river network to make flood routing possible. Finally, rainfall rates in this region are sufficiently high, and the soil sufficiently impermeable, to provide for the construction of runoff models that are not highly dependent on poorly resolved antecedent conditions.
Assuming that precipitation can be predicted with reasonable accuracy, this natural laboratory can be used to test the runoff and routing aspects of flood prediction, with the data sets applied to less severe events in order to develop better infiltration algorithms. At the same time, the calculated and measured surface precipitation fields and runoff hydrographs can be compared to those predicted by standard methods using the Denver or Cheyenne WSR-88D radars. This comparison already has been performed by Petersen et al. (1999) for the CHILL and Cheyenne WSR-88D radar-derived precipitation fields during the 1998 Spring Creek flood in Fort Collins, Colorado.

Once the hydrology of a flood has been studied, meso- and storm-scale models can be employed for addressing a variety of problems outlined in the previous section. If the hydrologic models perform sufficiently well, then it might be inferred that the flood prediction model is accurate enough to use the predicted flood hydrographs throughout the drainage basin, that is, to use them in the nongauged parts of the basin to more critically evaluate the details of the rainfall algorithms. Downscaling techniques can be similarly applied. Once the hydrologic model is known to be accurate under a wide variety of precipitation conditions, it can then be used with the gauge data to test the dominant features of the various precipitation algorithms even when CHILL or S-pol data are not available.

The Front Range natural laboratory described in the previous three paragraphs has most of the desired features of an effective research site. Predictable atmospheric conditions lead to a set of reasonably frequent, high-intensity events, the precipitation from which can be measured accurately with instruments that normally reside in the area. The nature of the river networks and land surface provides an opportunity for substantial simplifications for the multicomponent flood-predicting hydrologic model, and the river gauge network is sufficient to provide a set of good comparisons of predicted and measured flood hydrographs. Although the results of studies in this natural laboratory would not be directly transferable to other areas, the knowledge gained would be of general scientific value. For example, small watersheds and shallow soils are characteristic of the mountainous Hawaiian Islands, which also are subject to flash floods.

2) SITE 2: FLOODS CAUSED BY INTENSE CONVECTIVE RAINFALL THAT LANDS ON PRECONDITIONED GROUND

When heavy rain falls on more heavily vegetated terrain than discussed above, the condition of the soil and vegetation at the time of the event becomes important in modeling the runoff. This complicates considerably the hydrologic component of the flood problem. Careful investigation of this aspect of flood prediction is best done in a well-instrumented, regional-scale drainage basin such as the Walnut River watershed, which was selected for instrumentation in the Cooperative Atmospheric Surface Exchange Study (LeMone et al. 2000). The long-term instrumentation in and around this watershed includes 75 rain gauges (25 operational/climatological, 5 installed by Argonne National Laboratory, and 45 installed by Oregon State University); the Wichita, Kansas, WSR-88D radar located 20 km to the west; 5 operational stream gauges; and the Atmospheric Boundary Layer Experiments array (ABLE: LeMone et al. 2000), operated by Argonne. The ABLE array includes three profiling sites at the vertices of a 60-m equilateral triangle, each equipped with a surface meteorological station, a minisodar, and a 915-GHz wind profiler with radio acoustic sounding system. Within the triangle lies an operations center with a fourth profiler and meteorological station and three flux sites, providing complete surface energy budgets, along with soil temperature and moisture.

Accurate spatial structures of the temporally evolving precipitation field are essential, but not sufficient to generate an accurate flood prediction when heavy rain falls on heavily vegetated terrain. Rather, an accurate surface moisture model that accounts for the dominant atmospheric and biophysical processes also is necessary. Moreover, when applied on a regional scale, this surface condition model needs to provide a hydrologic feedback to the mesoscale model.

It is unlikely that soil moisture and other surface conditions ever will be measured directly and continuously in space and time. Instead, surface conditions will be sampled over several areas (e.g., 100 m × 100 m in area) selected by virtue of their surface and soil characteristics, plant cover, and location. For example, soil moisture could be sampled using time-domain reflectometry, for example, as in the Oklahoma Mesonet, with fluxes determined using the eddy correlation technique. However, sites and techniques would likely vary depending upon data source and acquisition capabilities. The measurements would be integrated with satellite and other meteorological data to obtain maps of surface skin temperature as well as heat and moisture fluxes. These, in turn, would be assimilated into and updated by a mesoscale model.
3) SITE 3: FLOODS PRODUCED BY RAIN FALLING ON SNOW-COVERED GROUND

Two ideal natural laboratories exist for this type of flood problem: the mountainous Pacific Northwest, where low-elevation wet snow is common throughout the winter and can be subjected to prolonged warm rain at any time, and the upper Mississippi River basin of the GCIP Large Scale Area (IGPO 1999). While the ground beneath the snow is sometimes frozen, it is more commonly saturated. In either case, the state or phase of moisture in the soil determines the antecedent hydrologic conditions, and as such determines the time scale of the associated runoff. The primary difficulty for this problem arises from the availability of the radar observations in complex terrain. This problem is offset somewhat by the increased correlation of weather with terrain and the associated enhancement in predictability.

4) SITE 4: FLOODS ASSOCIATED WITH TROPICAL AND EXTRATROPICAL CYCLONES

Significant flooding can occur in association with both tropical and extratropical cyclones, though the location of the latter is perhaps more highly variable. Owing to the large variability in the tracks of land-falling hurricanes, a natural laboratory that is fixed in space probably is not feasible. However, mobile instrumentation, including Doppler radar (Wurman et al. 1997), provides a viable means for collecting relevant data. Fortunately, WSR-88D-based rainfall rate fields for this type of event are reasonably accurate, and these storms usually affect large areas so that a large number of river gauges is available for comparing predicted and observed hydrographs. The antecedent land surface conditions most likely will have to be obtained from limited surface station data combined with satellite imagery.

5) SITE 5: FLOODS CAUSED BY HEAVY RAIN IN THE MONONGAHELA RIVER BASIN

During the next several years, the NWS Office of Hydrology will use the Monongahela River basin upstream from Pittsburgh, Pennsylvania, to develop new techniques and procedures for use by its River Forecast Centers. The Monongahela is a hydrologically interesting setting for studying floods because of the complex topography and geology of the Appalachian Plateau, the large annual precipitation, and the relatively low regional evaporation rates. Floods in the region appear to be particularly sensitive to antecedent moisture conditions; therefore, improvement in the spatial and temporal resolution of precipitation would yield a substantial improvement in antecedent soil moisture estimates.

Substantial benefit could be gained by treating this river basin, or parts thereof, as a natural laboratory and conducting additional scientific studies within it owing to the extensive NWS presence and support. This would, however, require adding both hydrological and meteorological instrumentation. Possibly one or more multiparameter research radars could be deployed to obtain accurate spatial precipitation structures throughout the period of greatest flooding probability. A major risk associated with the Monongahela is the possibility that rain storms with sufficient intensity to overcome the antecedent hydrologic conditions will not occur during the experiment period. In this situation, a major effort to characterize these antecedent conditions would become an important component of any such effort.

5. Concluding remarks

It is not possible for a report of this type to cover all the relevant issues, or to credit the many studies that have provided the strong foundation upon which the community now has an opportunity to build. The many significant advances made during the past few years in observing technology, computational and networking resources, and the fundamental understanding of meteorological and hydrological phenomena have set the stage for what could be the greatest advances in the history of meteorology and hydrology. Because floods are among the most destructive and societally disruptive of all natural phenomena, it is imperative that the recommendations made herein be given careful consideration and refined through specific action plans and research efforts. It is the sincere desire of the authors that this report serves to educate the hydrological and meteorological community about their differences, their commonality, and the opportunities resulting from both.

Acknowledgments. The authors gratefully acknowledge the leadership and support of R. Carbone in his role as USWRP lead scientist during the tenure of PDT 9, and to R. Gall in this same role during the final stages of this report. The authors also are grateful for contributions by W. Hooke, T. Graziano, J. Schaake, L. D. James, and D. Lettenmaier. The PDT 9 workshop was supported by NOAA Grant NA27GP0232, and by NSF Cooperative Agreement ATM92-09181.
References


