23HYDRO 1.4

Optimizing precipitation estimates using merged observations and model output: A case study in the California Sierra Nevada Mountains

Edward I. Tollerud[1], John A. McGinley[1], Steven L. Mullen[2,*],Tomislava Vukicevic[3], Huiling Yuan[3], Chungu Lu[4], and Isidora Jankov[4]

[1]*NOAA Earth System Research Laboratory (ESRL), Global Systems Division Boulder, Colorado USA*

[2]*Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona USA*

[3]*Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado at Boulder, and Earth System Research Laboratory (ESRL), Global Systems Division, Boulder, Colorado*

[4]*Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado USA, and Earth System Research Laboratory (ESRL), Global Systems Division, Boulder, Colorado USA*

1. Introduction

For water management, timely quantitative precipitation estimates (QPE) are critical. Unfortunately, there remain huge uncertainties with QPE that can interfere with water management decisions. These uncertainties are particularly large and damaging in mountainous regions where accurate point measurements at gage sites are sparse and instrumentation is often difficult to maintain, and radar observations are plagued by blockage problems. Under these conditions, there is significant potential value for high-resolution precipitation forecasts to serve a role as part of a blended product that can help to guide solely observation-driven QPE. However, suitable procedures must be devised to blend the forecasts and observations.

Here, we describe an effort to improve grid point specific QPE and basin-averaged water volume estimates by applying variationally-driven ensemble methods. These methods have been successfully employed in attempts to assimilate other quantities, but their application to precipitation has not been fully realized.

We explore the following questions:

1. Can explicit and meaningful precipitation error covariances be recovered for a specific event or series of events?
2. Can these error covariances be used explicitly to recover improved precipitation estimates as compared to withheld rain gauge sites, and provide some estimate of the expected analysis error covariance?
3. Are these precipitation estimates an improvement over existing estimates?

In Section 2 we formulate the optimal data assimilation technique used to produce QPE fields. Section 3 describes our methodology to apply the technique to specific events in the Northern California Sierra Nevada Mountains (specifically the basin of the American River northeast of Sacramento). The resulting fields for one extreme event are described in Section 4, including comparison with analyses from other sources and techniques. Section 5 summarizes our findings and suggests improvements to the technique and potential uses for the optimal QPE analyses.

2. Background

The basic formulation of the optimal data assimilation techniques that are used today in weather analysis is readily derived from general stochastic inverse problem theory under assumption that information about the state of weather is given by quantities contained in the NWP model forecast and in observations. In inverse problem theory the information from a model and observations is combined by a conjunction of the Gaussian probability density functions (pdf's) resulting in a new, joint posterior pdf. This conjunction is typically

[*]Presenting Author

presented as direct application of the Bayes theorem, and is written

(1) $p_{posterior} \sim \exp(-J(x_a, y))$

where x_a and y are modeled and measured stochastic Gaussian quantities, respectively, *cons* is some arbitrary constant and J is the well known cost function (Kalnay, 2004), which we define as

(2)

$$J(x_a, y) = \sum_i \frac{1}{2}\left[(h_i(x_a) - y_i)^T R^{-1}(h_i(x_a) - y_i)\right] + \frac{1}{2}(x_a - x_b)^T P_f^{-1}(x_a - x_b)$$

Since precipitation is generally not normally distributed it is necessary that we transformation our precipitation variables to log form. The assumption is that log x is normally distributed. Thus in our application, x_a is the analysis state vector containing the log of all precipitation estimates on the analysis grid, y_i is the log of observation of type i, which could be either a radar or rain gauge estimate. is necessary to ensure and $h_i(x_a)$ is the transformation from the analysis state space into observation space, referred to in our context as the "observational network operator". $R_{y_i}^{-1}$ is the inverse covariance matrix of logged observational errors for each observation type. In the first term in (2) it is implicitly assumed that the observation errors from different observation types are uncorrelated. x_b is the log of background precipitation vector of the same length (would be some deterministic estimate of precipitation from a model) as the analysis precipitation vector. P_f^{-1} is the inverse covariance matrix of the logged background precipitation errors derived from the 3-km ensemble. This matrix is the focus of our ensemble post processing. For a limited domain P_f may be fully recoverable.

When the observational operators h_i are linear and are combined into one operator they are conveniently denoted as H. The posterior solution in this case is expressed respectively in terms of mean and covariance as:

(3)

$$x_a = x_b + \left(HP_f H^T + R\right)^{-1} H^T R^{-1}(y - Hx_b)$$

$$P_a = (H^T R^{-1} H + P_f^{-1})^{-1}$$

3. Methodology

The approach described here involves an analytical test bed with known observed and forecast precipitation fields and live tests with actual 3-km data model runs. The analytical test bed will be used to assess the possibility of recovering explicit error covariances and applying equations (3) above. In this controlled environment an assessment can be made of the accuracy of the recovered analyzed precipitation and its error characteristics. For our purposes, background error covariances are determined using a hybrid scheme that involves a linear combination of both time-dependent (climatological) and time-independent (ensemble forecast-based) covariances. The relative influence of the two are determined via the settable coefficient alpha in the equation

(4) $P_f = a\,P_c + (1\text{-}a)\,P_e$

where P_c is the climatological covariace from many cases, described below and a is a weight that goes from 0 to 1. P_e is the ensemble derived covariance.

Climatological covariances are determined from a set of 95 cases (actually 6h time periods) during the IOPs for the first two HMT-ARB field exercises in 2005-6 and 2006-7. Currently, a matrix inversion including all observations is applied as the numerical solution to the optimal QPE. We are also investigating an alternative, more efficient solution technique that involves serial inversion in model space, with the assumption that observational errors are uncorrelated.

a. Applying the technique during the HMT ARB

To assess the capability of this technique to produce useful QPE fields during heavy to extreme precipitation episodes, we have applied it to several Intensive Operating Periods (IOPs) of the HMT ARB field exercise during 2005-2006. For a description of this exercise and tabulation of IOPs, see the HMT website at http://www.esrl.noaa.gov/psd/programs/2006/hmt/. This paper focuses on results from IOP 4 starting on Dec 29 2005 and extending through the first 6h period of January 2006. Fig. 1 shows a snapshot of gage-corrected radar estimates of precipitation (Stage IV), a WRF model forecast, and gage observations for one 6h period during IOP 4. It is clear that this period represents a

very significant precipitation event in Northern California, with large regions of the two domains receiving over 1.5 in of rainfall during the 6h time period in both the observations and the forecasts. In general, the forecasts and Stage IV estimates are similar and of comparable magnitude, but the forecasts reveal more detail than the smoother Stage IV estimates. Much of this detail is due to terrain variations in both the coastal ranges in the western half of the larger domain and the Sierras to the East. The roughly east-west oriented precipitation maxima in the forecasts (largely missing in the Stage IV plots) that appear prominently in the smaller ARB domain are likely also a result of interactions with terrain features, which include ridges and valleys that trend from west-southwest to east-northeast (Fig. 2). We note that this ability of models to incorporate the impact of terrain is one of the major reasons that we believe blended fields including forecasts can be useful for QPE.

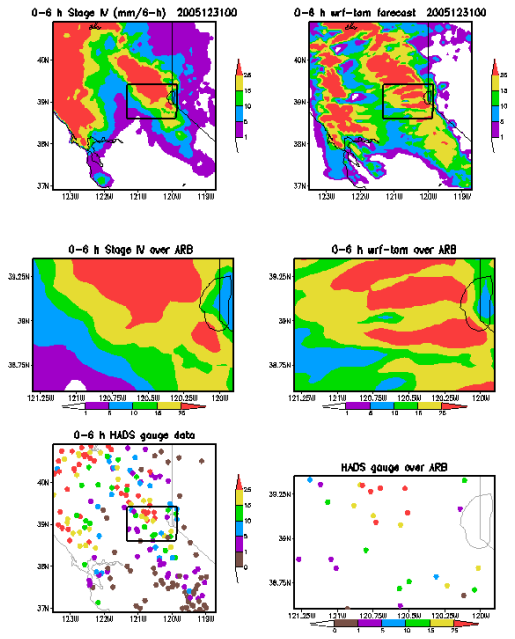

FIG. 1. Radar/Gage analysis for 6h period ending at 0000 UTC 31 December 2005 during IOP 4 (upper 2 left panels) and 6h lead time WRF forecast (upper 2 right panels) for the full 3-km WRF domain (upper row) and smaller ARB domain (middle row). Accompanying gage measurements from the opFerational hourly HADS network for the two domains are shown in the bottom row. All legends are in mm.

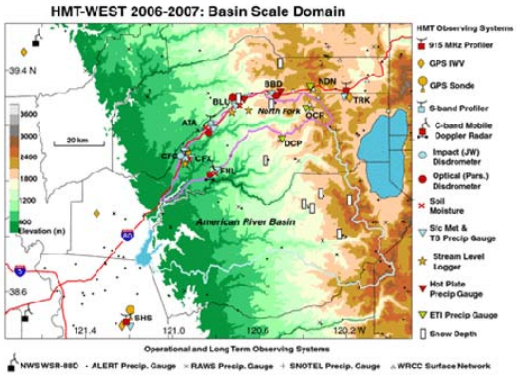

FIG. 2. Terrain features of the American River Basin. Locations of gages and other observation platforms are also shown. Image is taken from the HMT website http://www.esrl.noaa.gov/psd/programs/2006/hmt/.

b. WRF ensemble model specifications

Since 2005, NOAA/GSD/FAB has run a set of time-lagged multi-model ensembles for NOAA-HMT forecast applications (Yuan *et al.* 2008a). The time-lagged ensemble members in these runs are generated from a set of forecasts initialized every 6-hours using NOAA-LAPS, and evaluated at the same forecast projection time. Because the initializations of these forecasts are in a time-sequential fashion, these forecast ensembles contain important flow-dependent error information in the background fields (Lu *et al.* 2007). These time-lagged ensemble members can be generated via an on-line cycling along with model forecasts. Therefore, neither additional computation nor added cost are incurred. The retrieved background error statistics, including error variance and covariance, while capturing storm activity, remain largely balanced (Kim *et al.* 2008). This situation-dependent feature of background error covariances is crucial for a data assimilation system for short-range forecast applications. In the HMT application, we use 6h time-lagged ensembles to be consistent with model initialization and forecast configuration. Only the closest 3 lagged ensembles are used in order to avoid large error contamination due to longer model integrations. This limited ensemble size can cause serious rank-deficiency problem in the retrieved background error covariance matrix. In order to overcome this problem, a combination of time-lagged ensembles with a set of mixed physics ensembles are used to increase the ensemble size (Yuan *et al.* 2008b).

The mixed physics ensemble consisted of simulations performed by using Weather Research and Forecasting (WRF) numerical model with both Advanced Research and Forecasting (ARW) and Non-hydrostatic Mesoscale Model (NMM) dynamical cores. The integration domain covered a region of roughly 500x500km centered over central California (Fig. 1). For each event and for the WRF-ARW simulations the following microphysical schemes were used: Ferrier *et al.* (2002), Thompson *et al.* (2004) and Schultz (1995). The WRF-NMM simulations were performed by using Ferrier microphysics. For both dynamical cores and all microphysics the non-local mixing YSU PBL scheme (Noh *et al.* 2003) (an improved version of the MRF PBL scheme; Troen and Mahrt 1986) was used.

Fig. 3 provides a measure of model performance during IOP 4. On it are plotted a common precipitation verification metric, the equitable threat score (ETS) for ensemble and for a multi-scale analysis based completely on gage observations. The strategy used to produce adequate sampling (183 pairs) for meaningful scores is described in the next section, as is the selection of verification gages. Essentially, it amounts to aggregating statistics over eight 6h periods and several separate sets of withheld gages. For this episode of heavy precipitation, the most meaningful thresholds for the ETS comparison are those greater than .25 in. It is clear that the model forecasts perform well, obtaining larger (better) scores than the gage analysis at most of the thresholds.

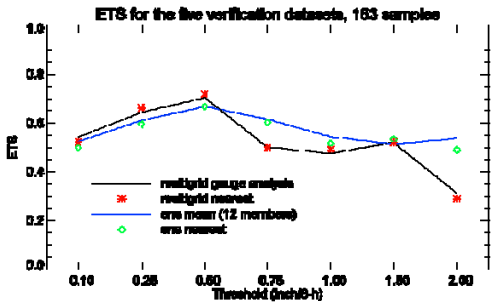

FIG. 3. Equitable threat scores (ETS) for two versions of the multi-grid gage analysis (STMAS; see text) and a WRF model ensemble mean forecast (ens) constructed from three lagged members and four model configurations. See text for a description of the selection of verification datasets. The continuous lines indicate interpolation to the exact location of verification gages; the points matched gage locations with the closest grid point.

c. Data and assessment strategy

As observational input to our data assimilation technique, we use 6h accumulated precipitation at operational hourly gages available from the Hydrometeorological Automated Data System (HADS) managed by the Office of Hydrology of the National Weather Service. These data are first screened by a set of automated daily quality control procedures (details available at http://www-frd.fsl.noaa.gov/mab/sdb/readme.txt). For the research diagnoses shown here the gage observations were then manually screened for possible errors not captured by the automated algorithms. However, given the difficulty of maintaining gages in good working order under the extreme conditions of terrain, elements, and heavy precipitation encountered during storms, it is likely that inaccuracies still exist in the gage measurements. The distribution of the full gage set in the ARB domain (Fig. 4, top middle panel) is not dense, but it is reasonably well distributed across the domain except in the east-central part and in the southwest corner. During each of eight 6h periods included in our analyses, the set of gages that survived both sets of screening numbered about 25.

To establish independent sets of gages for verification purposes, we have chosen to designate a set of verification gages for each of several separate analysis runs by withholding a small number of gages from each run. The locations of these gages are shown in Fig. 4; for the five runs shown, they number between three and five gages during each time period. This resulted in 183 verification pairs from the five initial runs and the eight time periods. Given the general similarity of the comparison analyses in the left column, which have been produced by the Space and Time Mesoscale Analysis System (STMAS, Xie et al. 2005), which applies a multi-scale analysis technique using only the gages shown, it appears that the analyses were not greatly impacted by the small number of withheld gages.

4. Results from IOP 4

A set of initial baseline results from IOP 4 are shown in Fig. 5. The optimal QPE field (hereafter referred to as "QPE") in this case has been constructed from all gages. Of the three, there is greater similarity between the ensemble mean forecast and the optimum QPE field, especially in the east-west oriented features. As

in previous discussion of Fig. 1, there appears to

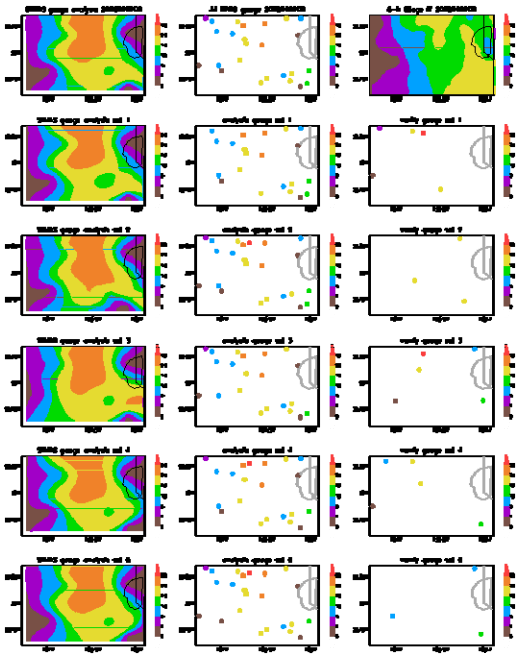

FIG. 4. Analyses and gage observations of accumulated precipitation for 6h period ending at 0000 UTC 1 January 2006 in the ARB domain. Left column displays multi-scale analyses (STMAS; see text for explanation) produced with all gages (top row) and five reduced gage sets. The gage observations used in the analyses are shown in the center column, and the gages withheld from the analyses for each of the five sets are in the right column. The upper right analysis displays the Stage IV radar/gage estimates for this period. Units on the color legends are in mm.

be greater detail (that is, more power at smaller spatial scales) in the forecast and QPE field than in Stage IV. In general, the Stage IV analysis is considerably smoother than the others. The optimum QPE also seems to have captured the finger of minimum precipitation (as indicated by gage observations) along the southern boundary of the domain better than either the Stage IV analyses or the WRF forecast. Another general impression is that the ensemble mean forecast overpredicts precipitation over much of the domain.

Two of the runs with withheld gages are shown in Fig. 6. As also appeared to be the case in Fig. 4 for the STMAS gage-only analyses at a later time, there is general similarity between these two runs in both the STMAS and QPE sets, a similarity that extends to the other three runs. A significant difference in smoothness and detail is also clearly seen between STMAS and the optimal QPE, which at this time has a striking east-west pattern. The STMAS analyses seem to retain larger values over the eastern domain boundary (the region including Lake Tahoe) than the optimal QPE.

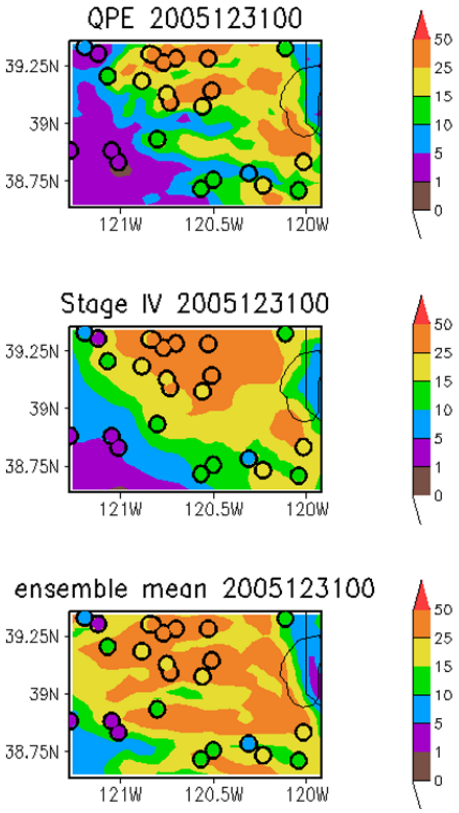

FIG. 5. Optimum precipitation analysis (QPE; top), Stage IV (middle), and WRF ensemble mean 6h forecast (bottom) valid at 0600 UTC December 31 2005 in the ARB domain. Color-coded circles are gages included in the QPE analysis. Legends are in units of mm.

While Fig. 6 displays a time period for which the optimal analysis performs well (at least qualitatively), other time periods and computational parameters can give worse results. For instance, a QPE plot during an earlier 6h period of IOP 4 when light general rain was falling over the domain (Fig. 7) exhibits large and apparently spurious precipitation maxima. Part of the problem in this case is likely traceable to a poor forecast; the ensemble mean is late getting the precipitation started, with the result that predicted rainfall is too small over most of the domain. This poor forecast in turn produces inaccurate error covariances and a poor optimal QPE field. Other factors may also be present, including the proper determination of weights for the climatological error covariances that are applied in the computations that lead to the QPE

analyses, and the best selection of lag times. We will discuss these possibilities in more detail in the conclusions. Although our preliminary sensitivity tests have not shown this to be universally true, it does appear that the optimal QPE computations encounter greater difficulty for scenarios with lighter and more general rainfall.

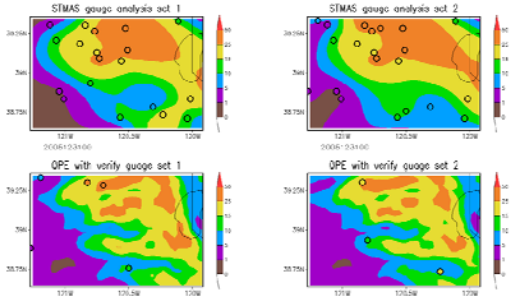

FIG. 6. Gage-only (STMAS; top row) analyses and optimal QPE (bottom row) for two sets of analyses of 6h precipitation ending at 0000 UTC 12 December 2005 in the ARB domain. Top row show locations of gages used in the analyses; bottom row displays gages withheld from analyses for later use in verification. Legends are in mm.

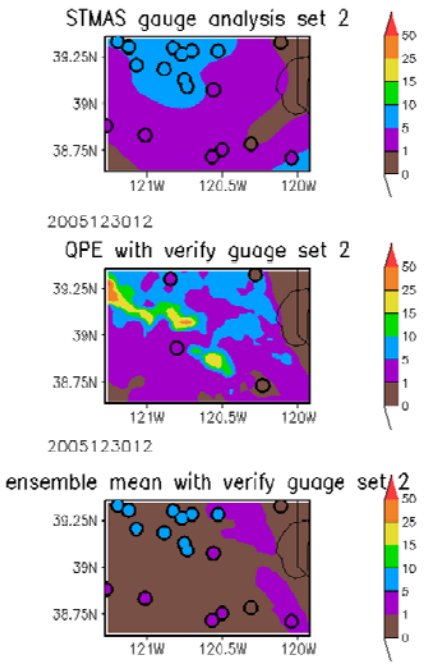

FIG. 7. STMAS and optimal QPE analyses (top and middle), and WRF ensemble mean forecast (bottom) for 6h precipitation ending at 1200 UTC 30 December 2005. Legends are in mm. Analyses gages are shown for the STMAS analysis and WRF forecast, and withheld gages are shown for the optimal QPE analysis.

With five sets of analyses during eight 6h periods, and 3-5 withheld gages for each, it is possible to compute quantitative statistics and verification scores with 183 verification gage precipitation observations matched with STMAS, optimal QPE, or ensemble forecast grid points. The scatter plot of Fig. 8 provides a general overview of the observation pairs that will make up these score computations. The forecast ensemble mean rainfall points show a strong tendency to appear above the one-to-one line of the verification observations, an indication of over-prediction. Both the STMAS and optimal QPE points generally straddle the line (that is, show less sign of overall bias), with the QPE points showing less scatter about that line than those for the STMAS analyses. There are a few QPE outliers that may account for some reduction in the correlation values; see for instance the QPE point with 50 mm precipitation that matches to a gage value of less than 10 mm. Overall, the ensemble forecast precipitation and the optimal QPE have correlation coefficients that are roughly equal, while the STMAS correlation is well below that of the other two.

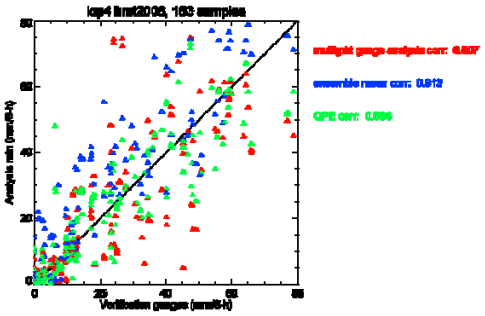

FIG. 8. Scatterplot of verification gage observations (horizontal axis, mm) against nearest grid point values of STMAS gage-only analyses (red), 12-member ensemble mean WRF forecasts (blue) and optimal QPE analyses (green) for 8 6h periods during IOP 4 (horizontal axis, mm). Spatial correlation values are also shown.

The IOP 4 domain-averaged precipitation rates and error estimates of Fig. 9 reveal, first, that optimal QPE and STMAS are indeed not biased as compare to overall verification gage averages. WRF forecasts, on the other hand, show a full 6 mm large bias over this IOP. While the mean absolute error between the ensemble mean and gages, and between STMAS and gages, is about the same, for QPE it is significantly less. Similarly, the root mean square error for QPE is also well below that of STMAS or WRF forecasts. This improved

performance for optimal QPE is reflected also in the equitable threat scores (ETS) of Fig. 10. For the range of thresholds most meaningful for this heavy rain event, (between 0.25 and 1.5 in), the QPE ETS is larger than (or in one case equivalent to) the other two, and occasionally is a full point better either of them.

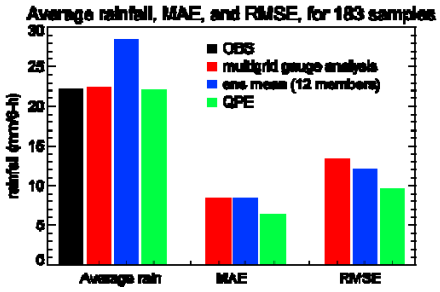

FIG. 9. ARB domain-average 6h rainfall (left bar cluster), mean absolute error (middle) and root-mean-square error (right) for all verification pairs during IOP 4. The black bar indicates straight average including all gages in the domain; other colors are as indicated in the legend.

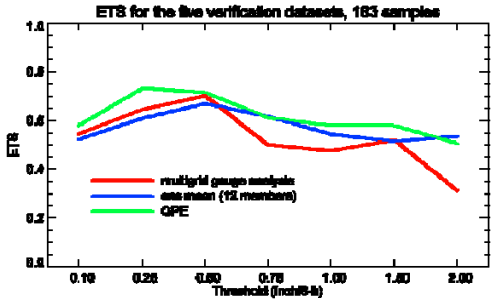

FIG. 10. Equitable threat scores (ETS) computed at verifying gage locations for the analyses and forecasts indicated.

Besides quantitative scoring, another assessment of comparative analysis value can be a determination of the detail revealed by the analysis. The four panels of Fig. 11 show the spatial power spectrum of the three analyses and the WRF ensemble mean forecast for one time period inside the ARB domain. The most striking features on the spectra are the peaks near 40 km, which we interpret as the general east-west upslope signal, and another near 12 km, which is roughly the distance between the east-west-oriented terrain ridges in the domain and between the analysis maxima noted previously in the ensemble mean forecasts and in the optimal QPE fields (Figs. 1 and 6). If the 12 km spectral peak is a reflection of these closely-space ridges, then it is clear that the forecast and the optimal QPE perform better at retaining the features of the precipitation field that result from terrain interactions.

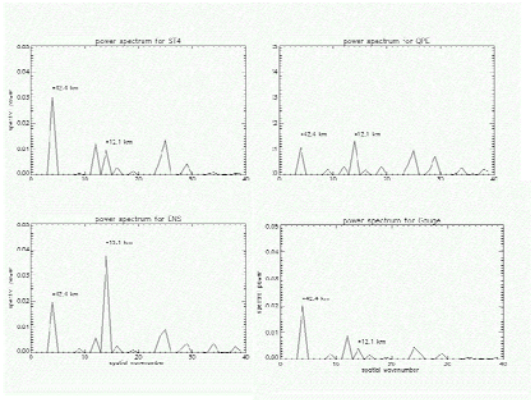

FIG. 11. Spectral estimates for precipitation totals for 6h period ending 0000 UTC 31 December 2005 during IOP4 for (clockwise from upper left): Stage IV precipitation estimates (ST4), optimum analysis (QPE), gage-only multi-scale analysis (Gage), and WRF ensemble mean (ENS).

5. Conclusions and Further Research

In general, this case study demonstrates that combining high-resolution ensemble forecasts with gage data using an optimal estimation methodology can successfully produce high-resolution QPE over a mountainous river basin that is superior to gage-only analyses (as measured by quantitative verification scores using withheld gages). However, it appears that occasionally bad error covariances from poor ensemble forecasts or other causes can result in spurious extreme rainfall in the QPE fields. These results suggest that further investigation of the forecast skill of ensemble forecasts is needed. Other important subjects for future research involve the specification of computational parameters in the optimal analysis methodology. For instance, sensitivity studies already performed demonstrate that when error covariances derived from scenario-specific model climatology are strongly weighted, the root mean square errors of the QPE fields are smaller (Fig. 12). The figure also shows that for low weightings, there is a dramatic dependence on the length of forecasts (i.e., number of lags). On the other hand, the QPE analysis is less sensitive to the inclusion of more time-lagged members when the model climatology weights are larger. The idealized parameter space diagram demonstrates this general movement toward better analyses with weighting increase

and forecast length decrease. More sensitivity tests of this kind are required before the best combination of parameters can be specified. An additional conclusion to be drawn is that careful quality control of gage data is critical for accurate QPE analyses.

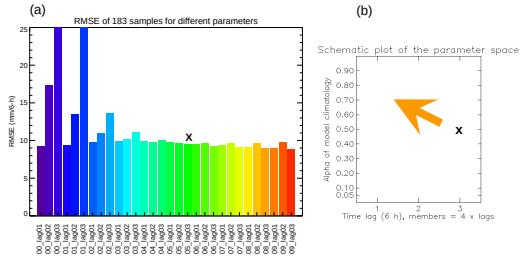

*FIG. 12. (a): Root-mean-square errors (RMSE) during IOP 4 for different combinations of model time lag ensemble members (1-3 6h periods) and climatological error covariance weightings (0.1 to 0.9; coefficient alpha in Section 3). (b): Idealized schematic parameter space diagram for model time lags and model climatology weighting. The '**X**' symbols on both panels indicate the set of parameters used in verification computations and analyses in previous figures; the arrow in panel (b) indicates the direction in parameter space required to move toward improved analyses.*

Other future plans include extension of this research to other IOPs during the HMT project and to a larger region including more river basins in the 3km model domain to capture a larger sample set of observations. Both verification scores and the spectral analysis results can be better confirmed by this improved sampling. Other data sources such as radar observations during the HMT project can be added to the QPE analysis. Possibly the inclusion of radar data can help confirm the validity of the small-scale ridge-valley circulations implied by the ensemble forecasts and the precipitation estimates, as well as by the spectral analyses. The QPE scheme will be examined for different ensemble configurations, such as combinations of various physics options and multiple models. With the present results in mind, a further question that can be addressed is this: If we assume a given QPE field is truth, is it possible to design a rain gage network that optimizes basin averages and narrows the analysis PDF in meaningful ways?

Acknowledgments

We thank Steve Koch, GSD director, for funding for this project through the Director's Discretionary Research funds. We also thank Annie Reiser for her technical review.

References

Albers, S., J. A. McGinley, D. Birkenheuer, and J. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analyses of clouds, precipitation, and temperature. *Wea. Forecasting*, **11**, 273-287.

Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and rainfall scheme in the NCEP Eta model. Preprints, *15th Conf. On Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 280-283.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103-120.

Kalnay E., 2004: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Kim, O.K., C. Lu, J. A. McGinley, and J.-H., Oh, 2008: Recovery of mesoscale error covariances using time-lagged ensembles. In preparation to submit to *Mon. Wea. Rev.*

Lu, C., H. Yuan,, B. Schwartz, and S. Benjamin, 2007: Short-range forecast using time-lagged ensembles. *Wea. Forecasting*, **22**, 580-595.

Noh, Y., W.-G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401-427.

Ralph, F. M., P. J. Neiman, and R. Rotunno, 2005: Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Mon. Wea. Rev.*, **133**, 889-910.

Schultz, P. J., 1995: An explicit cloud parameterization for optional numerical weather prediction. *Mon. Wea. Rev.*, **123**, 3331-3343.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. *Mon. Wea. Rev.*, **132**, 519-542.

Troen, I., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **47**, 129-148.

Xie, Y., S.E. Koch, J.A. McGinley, S. Albers, and N. Wang, 2005: A sequential variational analysis approach for mesoscale data assimilation. *21st Conf. on Weather Analysis and Forecasting*, Washington, D.C., Amer. Meteor. Soc., CD-ROM, 15B.7.

Yuan, H., J. A. McGinley, P. Schultz, C. Anderson, and C. Lu, 2008a: Evaluation and calibration of short-range PQPFs from time-phased and multimodel ensembles during the HMT-West-2006 campaign. *J. Hydrometeorology*, in press.

Yuan, H., C. Lu, J. McGinley, P. Schultz, L. Wharton, and B Jamison, 2008b: Short-range quantitative precipitation forecast (QPF) and probabilistic QPF using time-phased multi-model ensembles. In review, *Wea. Forecasting.*