9A.6 A HYDROLOGIC APPROACH TO EVALUATING QUANTITATIVE PRECIPITATION ESTIMATES

Jonathan J. Gourley[1] and Baxter E. Vieux[2]
[1]Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK
[2]University of Oklahoma, Norman, OK

1. INTRODUCTION

There exists a need to evaluate quantitative precipitation estimates (QPEs) from remote-sensing platforms such as radar and satellite in a hydrologically relevant way (Ciach and Krajewski 1999). Traditionally, QPEs under development are compared to rain gauges. Evaluation of QPEs solely by rain gauge can be useful but is lacking where sparsely populated gauge networks do not capture inherent rainfall variability. In addition, the scale difference between the orifice of a rain gauge and radar bins spans 8 orders of magnitude (Droegemeier et al. 2000). Lastly, multisensor algorithms are beginning to use the rain gauge estimates in their schemes, thus reducing the independence between the predicted and observed variables.

Recently, efforts have begun to evaluate the sensitivities of different rainfall estimates or forecasts on predicted hydrologic variables using hydrologic models. The accuracy of differing inputs may be difficult to assess when uncertainty in the model structure, parameters or even observations of output can dominate the overall prediction uncertainty. Moreover, model parameters can be calibrated to expect a given input (e.g., rain gauges) and even correct for systematic biases inherent in the rainfall estimates. Successive recalibration of the model is required for each input separately. In some cases, a long period of observations needed for model calibration is not available for new sensing platforms such as rainfall estimates from polarization diverse radars. The study reported herein develops a supplemental ensemble strategy to evaluate 9 different precipitation algorithms that are input to a physics-based, distributed hydrologic model independently. The “ground truth” in this case is the streamflow observed at the basin outlet.

2. METHODOLOGY

Hydrological evaluation is accomplished by an ensemble approach that takes into account parameter uncertainty and various multisensor QPEs. In this case, performance is viewed from the perspective of hydrologic model predictions compared with streamflow observations. The goal here is to determine which model input will yield the most accurate streamflow predictions in a probabilistic sense. It is of specific interest to the developers of QPE algorithms to know the accuracies and sensitivities of each product. For example, what is the impact of incorporating satellite data in the estimation scheme? Does this differ for events that are dominated by convection versus those that contain significant precipitation from stratiform clouds? Which gauge correction method works best, a mean field bias adjustment, a local bias adjustment, or both? Does the inclusion of gauge data in the products dramatically improve their accuracies? How does the model perform using gauge data as input alone? While this ensemble approach is demonstrated on a single basin, for specific precipitation estimates, using a single model structure, the generality of the method is not diminished.

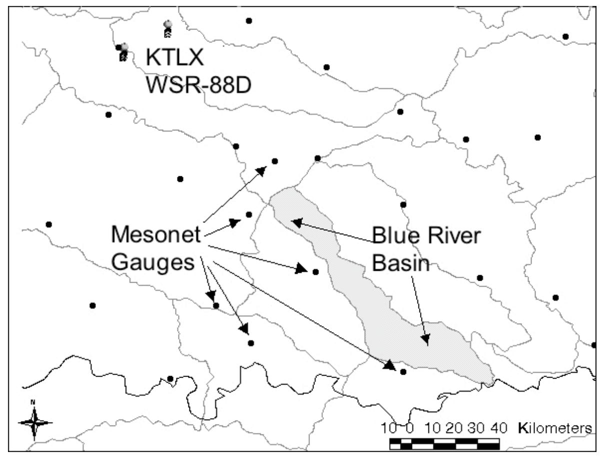

Fig. 1 - The Blue River study basin’s location (shaded in gray) relative to the nearest WSR-88D radar and Mesonet rain gauges.

2.1 *Blue River Basin*

An existing natural outdoor laboratory for hydrologic research is located in the Blue River Basin, near Blue, OK, U.S. (Fig. 1). The Blue River Basin drains about 1200 km^2. The headwaters of the basin are about 80 km away from the nearest weather radar, KTLX, while the basin outlet is over 200 km away from a radar. The long distance from radars makes this basin a good candidate for evaluating QPE algorithms at far range. The Blue River Basin is also attractive for hydrologic research due to the natural characteristics of the basin. There are no reservoirs in the basin, and there are very few known diversions at this time. The focus in this natural outdoor laboratory will be to demonstrate a new methodology in comparing different QPE algorithms by examining the basin response in a probabilistic framework.

2.2 *QPE SUMS precipitation estimates*

Quantitative Precipitation and Segregation Using Multiple Sensors (QPE SUMS; Gourley et al. 2001) is an automated precipitation algorithm that

* *Corresponding author address:* Jonathan J. Gourley, NSSL, 1313 Halley Circle, Norman, OK 73069; email: gourley@ou.edu

produces an ensemble of precipitation estimates having various levels of complexity. QPE SUMS incorporates radar, satellite, rain gauge, lightning, and model analyses in its estimation scheme to generate 9 different products. A brief explanation of the QPE SUMS precipitation ensemble members follows. The gauge-only analysis (GAG) utilizes 15-min rainfall accumulations from the Oklahoma Mesonet in addition to hourly rain gauge accumulations available from the US Geological Survey, National Weather Service, and other urban networks. The point data are analyzed to a common 1x1 km grid using a Barnes objective analysis scheme with the weighting function and cutoff radius set appropriately for the median gauge spacing. The radar-only product (RAD) relies on a mosaic of reflectivity data. This member is equipped with a convective-stratiform identifier that enables it to utilize different Z-R/Z-S equations based on the classification. The RAD product also gets the benefit of using lightning data to determine if there is convection at a given grid point. The RAD product is adjusted hourly using a mean field bias adjustment to yield the radar-mean field bias adjusted product (RAD-G). In addition to this adjustment, a "local bias" is computed at each gauge location by determining the difference between GAG and RAD accumulations. The local biases are objectively analyzed to the common grid using the same parameters used with the GAG product to produce a local bias field. The local bias field is then added back to the gridded RAD product to yield the radar-local bias adjusted product (RAD-LG). Another member (RAD-LGG) is produced by first removing the mean field bias from the RAD product and then applying the local bias adjustment scheme.

The multisensor product (MS) performs identically to the RAD product for grid cells that have been identified as convective. This member differs substantially from the RAD product for stratiform precipitation. In short, a scheme that is based on Gourley et al. (2002) adaptively calibrates infrared satellite data using radar-based precipitation rates that are measured *below the bright band*, and applies these satellite-based rainfall rates to grid points at far range from radar where it is sampling well above the melting layer. The goal of this technique is to reduce the range-dependence commonly associated with stratiform precipitation. The MS member is adjusted by gauges analogously to the RAD product described above to yield the multisensor-mean field bias adjusted (MS-G), multisensor-local bias adjusted (MS-LG), and multisensor-mean field bias removed and local bias adjusted (MS-LGG) products. All 9 QPE SUMS members are available on the same 1x1 km grid and have been aggregated so that they are available hourly.

2.3 Vflo™ distributed hydrologic model

The solution to the kinematic wave equation and modifications to account for trapezoidal channel routing in a GIS environment resulted in the model, *r.water.fea* (Vieux and Gauer 1994). The solution method solves the 1-D conservation of mass and momentum equations using finite elements in space and finite difference in time. Using this finite element approach a new model was developed called Vflo™, which may be used for event or continuous applications. Vieux and Vieux (2002) describe the deployment of the model for several watersheds including the Salt and Verde Watersheds in Arizona. Parameters used in the ensemble simulation study are initial soil moisture content, saturated hydraulic conductivity, and Manning roughness coefficient. Initial soil moisture content is treated as a single value over the basin, i.e., lumped, while the saturated hydraulic conductivity and Manning roughness coefficient parameters are spatially distributed. Following Vieux and Moreda (2003), these parameter maps are perturbed by multiplying them by scalars, thus preserving the spatial variability but varying the magnitude. Vflo™ has improved channel routing capabilities because flow is routed through measured cross-sections. Rating curves are used to represent the complex hydraulics in channels and over-bank flow in natural channels, e.g., the Blue River.

2.4 Evaluation using a model parameter ensemble

In environmental modeling, uncertainty exists with inputs, parameter values, model physics, and observations of the state, or output. Ensemble forecasting is based on the acknowledgment that the physical system being modeled may not be fully understood or observed, and thus it is not reasonable to supply deterministic predictions. Ensembles are used to predict the probability of future states by exploring several possibilities as completely as possible. A model parameter ensemble is used here to determine which model input produces an outcome that proves most similar to observations. This ensemble strategy avoids the need to calibrate the parameters for each input separately, store all the parameter settings, and then verify the calibrated, deterministic predictions on an independent data set. This probabilistic calibration method recognizes that there may not be a single minimum in the parameter response surfaces due to the uncertainties inherent in environmental modeling. However, there may be several different regions in the parameter space that result in equally acceptable simulations.

The 9 different QPE SUMS rainfall products are input to the Vflo™ model independently. Probability distributions are computed in this study using a 125-member ensemble by perturbing the soil moisture, saturated hydraulic conductivity, and Manning roughness coefficient parameters. The soil moisture is varied from 20% to 100% initial soil saturation in increments of 20%. As discussed in section 2.3, scalars are used to perturb the spatially distributed fields of saturated hydraulic conductivity and Manning roughness coefficient from their default values. The multipliers range from 0.25 to 1.75 in increments of 0.375. Simulations are then performed using a single model structure for each parameter combination so that the entire parameter space in the model has been explored. The time of maximum discharge (TIME), maximum discharge (PEAK), and the total volume of water normalized by the basin area (VOLUME) are derived from the observed and simulated hydrographs. The 125-member ensemble is then used to create discrete probability distributions of the 3 aforementioned hydrologic variables for each model input separately. Ten classes are used for each probability distribution where the minimum class and class interval is set to 25% of the observed PEAK and VOLUME values. The minimum class for the TIME variable is set to the observed TIME value minus 24 hours, and a class interval of 5 hours is used.

Verification of probability forecasts for multicategory events is accomplished using a ranked probability score (RPS). The RPS is sensitive to distance such that high probabilities assigned to events that are far displaced from the observed outcome are penalized more heavily than high probabilities assigned to events closer to the outcome. The RPS relies on the sum of squared errors so that values nearest to 0 indicate the best performance relative to observations. With 10 classes being used to compute the probabilities, the worst RPS score possible is 9. In the future, a resampling test will be employed to determine the significance levels.

3. RESULTS

The model parameter ensemble is used here to evaluate 9 different precipitation inputs for 3 hydrologic events that occurred on the Blue River. For the first event, precipitation began shortly after 1200 UTC 23 October 2002 and ended at 0300 UTC 25 October 2002 resulting in a PEAK of 56.7 cms at a TIME of 0000 UTC 26 October 2002. The VOLUME was 6.1 mm. The second and third events occurred in October and December 2002 and were lighter than the first event having PEAK values of 16.4 and 13.4 cms and VOLUMES of 2.3 and 2.5 mm, respectively. It should be noted that some data outages occurred during the second and third cases resulting in the loss of 4 estimators.

For flash flood forecasting, it is important for a model to be able to match the TIME, PEAK, and VOLUME of a hydrograph. The TIME indicates when a flood wave may impact a region of interest, while the PEAK and VOLUME variables are related to the magnitude of flooding. For the purposes of QPE development, the VOLUME is the most informative variable. Recall the outlet of the Blue River Basin is 200 km from a nearby radar, KTLX. The extent at which each product may possess biases, especially at far range, is important in assessing the relative strengths and weaknesses of a particular QPE. These biases will be revealed most explicitly upon analysis of the VOLUME results.

Tables 1-3 list the RPS values for the TIME, PEAK, and VOLUME with the averaged RPS values summarized in the last column. The RAD ensemble member has the lowest overall RPS value (indicated in boldface) for the TIME variable. This isn't too surprising because radar data have the highest temporal and spatial resolution and therefore should match the timing of the flood crest the best. The MS and gauge-biased algorithms incorporate data that have coarser temporal resolutions which evidently affects the timing of the predicted maximum discharge. Table 2 shows that the MS-G product is capable of predicting the most accurate peak discharge for the events studied. Notice how the MS product outperforms the RAD product, and both of these products benefit from a mean field bias adjustment. However, the local bias adjusted products (with and without removal of mean field bias) offer no

TABLE 1. Ranked probability scores for the time at which the maximum discharge occurred.

QPE product	Case 1	Case 2	Case 3	AVG
GAG	1.10	3.05	0.94	1.70
RAD	1.02		0.65	**0.84**
RAD-G	0.32		1.68	1.00
RAD-LG	1.14		0.93	1.04
RAD-LGG	0.99		0.88	0.94
MS	2.14	3.65		2.90
MS-G	1.79	2.13		1.96
MS-LG	1.10	3.05		2.08
MS-LGG	1.04	2.11		1.58

TABLE 2. Ranked probability scores for the maximum discharge.

QPE product	Case 1	Case 2	Case 3	AVG
GAG	0.44	4.28	5.64	3.45
RAD	0.43		5.20	2.82
RAD-G	1.87		0.89	1.38
RAD-LG	0.45		5.64	3.05
RAD-LGG	2.42		5.54	3.98
MS	1.21	1.25		1.23
MS-G	0.43	0.69		**0.56**
MS-LG	0.44	4.30		2.37
MS-LGG	0.78	4.88		2.83

TABLE 3. Ranked probability scores for the flow volume.

QPE product	Case 1	Case 2	Case 3	AVG
GAG	1.86	5.14	5.91	4.30
RAD	1.88		5.94	3.91
RAD-G	4.11		2.66	3.39
RAD-LG	1.86		5.85	3.86
RAD-LGG	4.14		5.85	5.00
MS	0.51	0.70		**0.61**
MS-G	1.65	2.19		1.92
MS-LG	1.86	4.82		3.34
MS-LGG	2.58	5.97		4.28

improvements and perform more similarly to the GAG product. Figure 1 shows the proximity of rain gauges around the Blue Basin. The density of rain gauges in this location causes the local bias adjusted products to place a high degree of weighting on the nearby rain gauges. If these rain gauges are not representative of the areal rainfall amounts in the basin, then errors can actually be introduced in the local bias adjusted products. Table 3 shows the QPE estimator that is able to predict the total volume of water that is discharged at the basin outlet most accurately is the MS product. For the VOLUME variable, the inclusion of rain gauge data and successive adjustment techniques offer no improvements to the MS product. The RAD product, on the other hand, does realize modest improvements with a mean field bias adjustment and to a lesser degree a local bias adjustment. For the PEAK and VOLUME variables, the local bias adjustment with mean field bias removed substantially reduces the MS and RAD product accuracy and essentially reproduces the gauge-only analysis.

4. DISCUSSION

This paper has developed a 125-member model parameter ensemble by varying 3 sensitive parameters identified in the Vflo™ distributed parameter model. Precipitation inputs are supplied to the hydrologic model from the QPE SUMS precipitation algorithm. This algorithm is currently capable of producing 9 different precipitation products, with each using various sensors including radar, satellite, lightning, model output, and rain gauges. Each product is input to the model independently and an ensemble of hydrologic predictions is produced. The skill of each initial condition is assessed by comparing simulations with observations of streamflow and computing a ranked probability skill score from the probability distributions.

The results from 3 hydrologic events on the Blue River indicate that a mean field bias adjustment to the RAD and MS products improves the model skill for the PEAK and VOLUME variables. One notable exception is that the MS algorithm possesses the most skill for the VOLUME variable without *any* adjustment from rain gauges. The non-adjusted MS product was designed to reduce the range-dependency of precipitation estimates for stratiform precipitation by introducing calibrated satellite data in its scheme. From this initial investigation, it appears that these biases have been successfully mitigated with the MS product. It is also interesting to note that the local bias adjusted products have little skill and produce hydrologic simulations similar to the gauge-only analysis. Evidently, the point gauge accumulations do not represent the areal rainfall accumulations or local biases for the events studied.

The model parameter ensemble offers a new perspective in evaluating the accuracy of initial conditions. In this case, useful information about the inclusion of satellite data and rain gauge adjustment strategies in QPE products has been made available to the developers. Future studies will explore the relative skill of a model parameter ensemble versus an initial condition ensemble in hydrologic modeling. An initial condition ensemble is made possible through the use of the 9 members available from the QPE SUMS precipitation algorithm. Ultimately, an ensemble that incorporates the uncertainty in the rainfall estimates and in the model parameters may be necessary to accurately portray the total prediction uncertainty.

5. ACKNOWLEDGEMENTS

Funding for this research was provided under NOAA-OU Cooperative Agreement #NA17RJ1227. Support from the US Department of Education's Graduate Assistantship in Areas of National Need provided for the registration, publication, and travel costs.

6. REFERENCES

Ciach, G. J. and W. F. Krajewski, 1999: Conceptualization of radar-raingage comparisons under observational uncertainties. *J. Appl. Meteor.*, **38,** 1519-1525.

Droegemeier, K. K. and co-authors, 2000: Hydrological aspects of weather prediction and flood warnings: Report on the ninth prospectus development team of the U. S. Weather Research Program, *Bull. Amer. Meteor. Soc.*, **81**, 2665-2680.

Gourley, J. J., J. Zhang, R. A. Maddox, C. M. Calvert, and K. W. Howard, 2001: A real-time precipitation monitoring algorithm - Quantitative Precipitation Estimation and Segregation Using Multiple Sensors (QPE SUMS). Preprints *Symp. on Precipitation Extremes: Prediction, Impacts, and Responses*, Albuquerque, Amer. Meteor. Soc., 57-60.

Gourley, J. J., R. A. Maddox, K. W. Howard and D. W. Burgess, 2002: An exploratory multisensor technique for quantitative estimation of stratiform rainfall. *J. Hydrometeor.*, **3**, 166-180.

Vieux, B. E. and N. Gauer, 1994: Finite Element Modeling of Storm Water Runoff Using GRASS GIS, *Microcomputers in Civil Engineering,* **9**, 263-270.

Vieux, B. E. and F. Moreda, 2003 : Ordered physics-based parameter adjustment of a distributed model. In: *Calibration of Watershed Models*, Eds. Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., & Turcotte, R. American Geophysical Union, Washington, D.C., USA, p. 267-281.

Vieux, B. E. and J. E. Vieux, 2002: Vflo™: A real-time distributed hydrologic model. *Proceedings of the 2nd Federal Interagency Hydrologic Modeling Conference*, Las Vegas, Nevada.