UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CHARACTERIZING COMPONENTS OF UNCERTAINTY IN HYDROLOGIC MODELING USING AN ENSEMBLE APPROACH

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

JONATHAN JOSEPH GOURLEY

Norman, Oklahoma

2003

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/679

UMI Number: 3109074

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 3109074

Copyright 2004 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

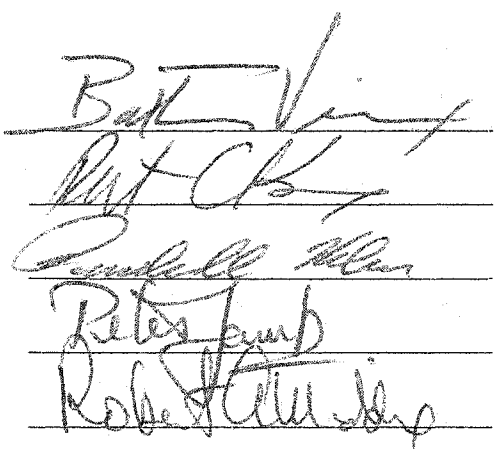
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright by JONATHAN JOSEPH GOURLEY 2003
All Rights Reserved.

CHARACTERIZING COMPONENTS OF UNCERTAINTY IN HYDROLOGIC MODELING USING AN ENSEMBLE APPROACH

A Dissertation APPROVED FOR THE
SCHOOL OF CIVIL ENGINEERING AND ENVIRONMENTAL SCIENCE

BY



ACKNOWLEDGMENTS

Funding for this study was made available by the Department of Education's Graduate Assistance in Areas of National Need Program as well as funding under NOAA-OU Cooperative Agreement #NA17RJ1227. Their support is gratefully acknowledged. Oklahoma Mesonet data are provided courtesy of the Oklahoma Mesonet, a cooperative venture between Oklahoma State University and The University of Oklahoma and supported by the taxpayers of Oklahoma.

Dr. Baxter Vieux supplied many meaningful insights regarding the specific methodology and hydrologic modeling in general. He has also been instrumental in getting me in tune with the hydrologic community. I would also like to thank Jean Vieux and the many employees at Vieux and Associates, Inc. who answered many questions about the *Vflo* ™ model and provided invaluable support for this effort. Barrett Stanke, a visiting undergraduate student working under the Research Experience for Undergraduates Program, helped out with a statistical study regarding climatological rainfall and runoff relationships. His assistance is appreciated. Also, Drs. Dave Stensrud, Kim Elmore, David Bright, Harold Brooks, Mike Baldwin, and Matt Wandishin, all employed at the National Severe Storms Laboratory, provided extremely useful information to me about statistics and ensemble modeling. This study would not have been successful without their guidance. I would like to thank members of the WISH team, NSSL

management, and close colleagues for being patient with me during the fall of 2003. Also, thanks to NSSL for providing computer resources and facilities needed for this study. I appreciate Drs. Bob Maddox, Bob Knox, Randall Kolar and Pete Lamb for serving on my committee. Their suggestions were quite helpful.

This dissertation would not have been possible without continuing moral support and sacrifices taken by my family. My parents are gratefully appreciated for encouraging me to set high standards, through direct interaction and by example, and to follow through in accomplishing those goals. They will always be regarded in the highest esteem by me. My wife, Karen, has also been very gracious to take over many responsibilities at home over the past couple years. Her love and support is valued the most. Lastly, although I doubt he can read this, I would like to acknowledge the unconditional support provided by my dog, Charlie. He has actually been by my side for a majority of the writing both at home and at work. He has been patient and thoroughly enjoyed the few moments I did take to run him or play with him outdoors. I owe Karen and Charlie both a long camping/fishing trip out West.

TABLE OF CONTENTS

	PAGE
Acknowledgements	iv
List of Tables	viii
List of Illustrations	x
Abstract	xvii
Chapter I. Introduction	1
Chapter II. Background	15
Section 2.1 Quantitative Precipitation Estimation	15
Section 2.1.1 Rain Gauges	15
Section 2.1.2 Radar	18
Section 2.1.3 Satellite	25
Section 2.1.4 Multisensor	28
Section 2.2 Hydrologic Models	35
Section 2.3 Model Calibration	48
Section 2.4 QPE Evaluation Methodologies	53
Section 2.4.1 Point Measurements	53
Section 2.4.2 Hydrologic Simulations	54
Chapter III. Methodology	57
Section 3.1 Research Objectives	57
Section 3.2 Experimental Plan	62
Section 3.2.1 Case Data	62
Section 3.2.2 Hydrologic Evaluation Using A Model Parameter Ensemble Approach	65
Section 3.2.3 Evaluation of the Characteristics of Uncertainty	82
Section 3.2.3.1 Propagation Characteristics of Input Uncertainty	82
Section 3.2.3.2 Identification of Uncertainty in Model Physics	84
Section 3.2.3.3 Estimation of Combined Prediction Uncertainty	89
Section 3.2.3.4 Optimum Number of Ensemble Members	91
Chapter IV. Results	96
Section 4.1 Hydrologic Evaluation of Quantitative Precipitation Inputs	96

Section 4.1.1 23 October 2002 Case 96
Section 4.1.2 28 October 2002 Case 117
Section 4.1.3 03 December 2002 Case 133
Section 4.1.4 Summary of Hydrologic Evaluation 148
Section 4.2 Evaluation of the Characteristics of Uncertainty 156
Section 4.2.1 Propagation of Uncertainty from Inputs to Hydrologic Predictions 158
Section 4.2.2 Identification of Uncertainty in Model Structure 173
Section 4.2.3 Estimation of Total Prediction Uncertainty 188
Section 4.2.4 Optimum Number of Ensemble Members 201
Chapter V. Summary and Conclusions 206
Appendix 216
Literature Cited 221

LIST OF TABLES

TABLE		PAGE
3.1.	A summary of the hydrologic events studied. The Observed Time refers to the time of peak discharge, the Observed Peak is the magnitude of the peak discharge, and the Observed Volume is the time-integrated discharge normalized by the basin area. See Chapter II, section 2.1.4 for a complete description of the rainfall products.	94
4.1.	Statistical evaluation of hydrologic predictions for the 23 October 2002 case. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). See Chapter II, section 2.1.4 for complete descriptions of rainfall inputs (in first column) and Chapter III, section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.	103
4.2.	Significance levels of RPS differences for the 23 October 2002 case. Significance levels are determined using a resampling test (see Chapter III, section 3.2.2 for complete discussion).	105
4.3.	As in Table 4.1 but for statistical evaluation of hydrologic predictions for the 28 October 2002 case.	121
4.4.	As in Table 4.2 but for significance levels of RPS differences for the 28 October 2002 case.	122
4.5.	As in Table 4.1 but for statistical evaluation of hydrologic predictions for the 03 December 2002 case.	137
4.6.	As in Table 4.2 but for significance levels of RPS differences for the 03 December 2002 case.	138
4.7.	As in Table 4.1 but for average statistical evaluation of hydrologic predictions for all three cases combined.	150

4.8.	Statistical evaluation of hydrologic predictions for the 12 November 1994 case. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). The rainfall inputs are perturbed by scalars as noted in first column. See section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.	160
4.9.	As in Table 4.2 but for significance levels of RPS differences for the 12 November 1994 case.	161
4.10.	Statistical evaluation of hydrologic predictions from combined input-parameter ensembles for cases listed in the left column. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). See section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.	176

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1.1.	The study area including the Blue River Basin (shaded in yellow), WSR-88D radars, Mesonet rain gauge locations (red filled circles), USGS gauging sites (light blue filled triangles), and a shaded relief map of Oklahoma.	13
3.1.	A schematic of the components of a simple model. Uncertainty may be present in the areas noted.	60
3.2.	A schematic of a typical hydrograph and the derived variables that are used in this study. The Time (sec) variable is the time at which the maximum discharge occurs, Peak (cms) is the magnitude of the maximum discharge, and Volume is the time-integrated flow under the hydrograph that's been normalized by the basin area (mm).	72
4.1.	Storm total precipitation plots for the 23 October 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c) radar with mean field bias removed, (d) radar with local bias adjustment, (e) radar with mean field bias removed and local bias adjustment, (f) multisensor, (g) multisensor with mean field bias removed, (h) multisensor with local bias adjustment, and (i) multisensor with mean field bias removed and local bias adjustment. A complete description of each algorithm is provided in Chapter II, section 2.1.4. Values are in inches as indicated in the color bar. Red dots correspond to Mesonet rain gauging locations and cyan triangles are USGS streamgauges.	99
4.2.	Observed streamflow at the Blue River, near Blue, OK gauging site (USGS #07332500) for the 23 October 2002 case.	102

4.3.	Probability density functions of the predicted time of maximum discharge (Time) for the 23 October 2002 case. The color coding of the curves (noted in the legend) corresponds to the different model inputs used to construct the model parameter ensembles. Observed Time is 0000 UTC 26 October 2002.	109
4.4.	As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak). Observed Peak is 56.7 cubic meters per second (cms).	110
4.5.	As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area. Observed Volume is 6.1 mm.	111
4.6.	Simulation bounds of the predicted time of maximum discharge (Time) for the 23 October 2002 case. The open boxes refer to the 50% quantile (median), while the bars correspond to the 5% and 95% quantiles. The horizontal double line is the observed Time.	114
4.7.	As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 23 October 2002 case.	115
4.8.	As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 23 October 2002 case.	116
4.9.	As in Fig. 4.1 but for storm total precipitation plots for the 28 October 2002 case from the QPESUMS products: (a) gauge-only, (b) multisensor, (c) multisensor with mean field bias removed, (d) multisensor with local bias adjustment, and (e) multisensor with mean field bias removed and local bias adjustment.	119
4.10.	As in Fig. 4.2 but for observed hydrograph for 28 October 2002 case.	120
4.11.	As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 28 October 2002 case. Observed Time is 0600 UTC 30 October 2002.	125

4.12.	As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case. Observed Peak is 16.4 cubic meters per second (cms).	126
4.13.	As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 28 October 2002 case. Observed Volume is 2.3 mm.	127
4.14.	As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 28 October 2002 case.	129
4.15.	As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case.	130
4.16.	As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 28 October 2002 case.	131
4.17.	As in Fig. 4.1 but for storm total precipitation plots for the 03 December 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c) radar with mean field bias removed, (d) radar with local bias adjustment, and (e) radar with mean field bias removed and local bias adjustment.	134
4.18.	As in Fig. 4.2 but for observed hydrograph for 03 December 2002 case.	135
4.19.	As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 03 December 2002 case. Observed Time is 1600 UTC 05 December 2002.	140
4.20.	As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case. Observed Peak is 13.4 cubic meters per second (cms).	141
4.21.	As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 03 December 2002 case. Observed Volume is 2.5 mm.	142

4.22.	As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 03 December 2002 case.	144
4.23.	As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case.	145
4.24.	As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 03 December 2002 case.	146
4.25.	As in Fig. 4.2 but for observed hydrograph for 12 November 1994 case.	159
4.26.	As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 12 November 1994 case. Observed Time is 2000 UTC 15 November 1994.	163
4.27.	As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 12 November 1994 case. Observed Peak is 215.2 cubic meters per second (cms).	164
4.28.	As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 12 November 1994 case. Observed Volume is 41.1 mm.	165
4.29.	As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.	167
4.30.	As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.	168

4.31.	As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.	169
4.32.	As in Fig. 4.1 but for storm total precipitation plots for the 25 August 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c) radar with mean field bias removed, (d) radar with local bias adjustment, (e) multisensor, (f) multisensor with mean field bias removed, and (g) multisensor with local bias adjustment.	175
4.33.	As in Fig. 4.2 but for observed hydrograph for 25 August 2002 case.	177
4.34.	Probability density function of the predicted time of maximum discharge (Time) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Time is 0900 UTC 28 August 2002.	179
4.35.	Probability density function of the predicted magnitude of maximum discharge (Peak) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Peak is 6.2 cubic meters per second (cms).	180
4.36.	Probability density function of the predicted time integrated discharge volume normalized by the basin area (Volume) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Volume is 1.6 mm.	181
4.37.	Scatter plot of predicted time of maximum discharge (Time) versus observed Time for the cases listed in Table 4.10. The open boxes refer to the 50% quantile (median), while the bars correspond to the 5% and 95% quantiles. The diagonal double line is the 1:1 line.	182
4.38.	As in Fig. 4.37 but for scatter plot of predicted magnitude of maximum discharge (Peak) versus observed Peak for the cases listed in Table 4.10.	183

4.39.	As in Fig. 4.37 but for scatter plot of predicted time integrated discharge volume normalized by the basin area (Volume) versus observed Volume for the cases listed in Table 4.10.	184
4.40.	Climatological trend of monthly rainfall-runoff relationships and soil moisture conditions on the Blue River Basin. The slope (blue curve) indicates the relationship between rainfall and runoff using 61 years of observations. The other curves (see legend) are monthly fractional water content values at depths of 5, 25, 60, and 75 cm. 7 years of soil moisture conditions from the Durant Mesonet site are used to construct the fractional water content climatologies.	186
4.41.	Probability density function of the predicted time of maximum discharge (Time) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Time is 0000 UTC 26 October 2002.	190
4.42.	Probability density function of the predicted magnitude of maximum discharge (Peak) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Peak 56.7 cubic meters per second (cms).	191
4.43.	Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Volume is 6.1 mm.	192
4.44.	Probability density function of the predicted time of maximum discharge (Time) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Time is 0600 UTC 30 October 2002.	194
4.45.	Probability density function of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Peak 16.4 cubic meters per second (cms).	195

4.46.	Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Volume is 2.3 mm.	196
4.47.	Probability density function of the predicted time of maximum discharge (Time) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Time is 1600 UTC 05 December 2002.	197
4.48.	Probability density function of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Peak 13.4 cubic meters per second (cms).	198
4.49.	Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Volume is 2.5 mm.	199
4.50.	Ranked probability score plotted as a function of the number of members comprising the ensemble for the 23 October 2002 case. The three curves correspond to predictions of time of maximum discharge (Time; open squares), magnitude of peak discharge (Peak; filled circles), and time-integrated discharge volume normalized by the basin area (Volume; open triangles).	203
4.51.	As in Fig. 4.50 but for the 28 October 2002 case.	204
4.52.	As in Fig. 4.50 but for the 03 December 2002 case.	205
A.1.	Absolute difference between predicted and observed peak discharge (a; in cms) and time-integrated discharge volume (b; in mm) for the 23 October 2002 case. The error is an average value considering different initial soil saturation settings.	217
A.2.	As in Fig. A.1 but for the 28 October 2002 case.	219
A.3.	As in Fig. A.1 but for the 03 December 2002 case.	220

ABSTRACT

In hydrologic modeling, uncertainties are known to reside in model inputs, i.e., rainfall estimates, model parameters, observations of streamflow, and in some cases in the model structure itself. Estimation of the total prediction uncertainty for a hydrologic forecast first requires knowledge of the error characteristics of input rainfall estimates. Traditionally, evaluation of quantitative precipitation estimates (QPEs) has been accomplished by comparing remotely sensed rainfall to point rain gauge observations. In addition to errors associated with rain gauge measurements, it has been noted that sampling sizes between a typical radar pixel and a rain gauge orifice differ by about eight orders of magnitude (Droegemeier et al. 2000). It is thus highly desirable to design an objective, quantitative methodology that evaluates the skill of precipitation algorithms at the hydrologic scale of application, a watershed. QPEs from different algorithms are input to the *Vflo* ™ hydrologic model. Thousands of simulations are performed in an ensemble fashion in order to "expose" each rainfall input to the entire parameter space. Probabilistic statistics are utilized to compare the predicted probability density functions (pdfs) to observations of streamflow. Results indicate the spatial variability of rainfall observed by radar is indeed important for skillful hydrologic predictions.

The developed ensemble approach is also used to evaluate the propagation characteristics of error in rainfall estimates to hydrologic predictions. Predictions of peak discharge and time-integrated discharge volume are shown to be very sensitive to rainfall perturbations. Ensembles are then constructed to include the combined uncertainty in QPEs and model parameters. Several case studies are utilized to show how the total prediction uncertainty can be accurately estimated. Moreover, additional sources of uncertainty are identified for a case where simulation bounds derived from predicted pdfs do not replicate observed behavior during summer months. This may be due to inadequate parameterization (e.g., initial abstraction) or to model structure. Soil moisture observations from the Oklahoma Mesonet are introduced to reveal some explanation for conditions where the Green and Ampt submodel may not adequately characterize the infiltration behavior during summertime low flows.

CHAPTER I. INTRODUCTION

Some of the worst natural disasters in the world have been a result of excess water or lack of it. Excess water can initially present itself in the liquid or frozen phase. These meteorological extremes, e.g. heavy rainfall or snowfall, become a hydrologic problem given facilitating boundary and initial conditions in the atmosphere and land surface. Predicting water transport and fate encompasses many physical aspects of the hydrologic cycle. Quantitative precipitation estimation/forecasting (QPE/QPF) is the first step in the hydrologic prediction process. Accurate depiction of the rainfall field does not necessarily result in accurate forecasts of streamflow; it simply provides a characterization of the initial state of the system. The accuracy of future states (i.e., predictions) is based on our knowledge of the system's dynamic behavior (i.e., the model) and how well we parameterize these known physical processes to simulate observed system behavior. The relationship between rainfall and runoff has been found to be nonlinear and must be modeled (Droegemeier et al. 2000). The study undertaken examines predictability issues associated with a hydrologic prediction system in response to differing precipitation inputs.

Accurate hydrologic prediction will impact many agencies involved in the protection of lives, property, and the natural environment. At small time and space scales, real-time streamflow forecasts can be used to warn against

hazardous runoffs potentially resulting in flooding. Floods and flash floods kill more people annually than any other weather-related disaster. On average, about 130 people in the U.S. lose their lives to floods in a given year. A majority of those deaths occur inside vehicles. In addition, property damages due to flooding average about $1 billion a year with occasional losses exceeding $5 billion. A real-time hydrologic prediction system may be used to warn the public and invoke mitigation strategies well in advance of the flood impact. A more recent threat to the U.S. is potential contamination of municipal water supplies by terrorists. It is believed that the hydrologic prediction system discussed herein can be coupled to a water quality model that models additional physical processes such as advection, dispersion and decay. This application can provide predictions as to the severity and extent of a toxic spill and the locations that may be impacted downstream. The hydrologic prediction system used in this study may also be coupled to a storm surge model. It can supply needed initial and boundary conditions as to the amount of discharge impacting inland estuaries. The storm surge model would then predict the additional volume of water impinging upon the estuary from the ocean. With this, flood inundation regions from a land-falling hurricane could be mapped out in advance so that mitigation strategies can be implemented. At longer times and larger spatial extents, snow pack on watersheds can be monitored and subsequent streamflow forecasts can be utilized in watershed management, like dam operations and for agricultural purposes.

Overall, the hydrologic prediction system serves as a necessary component to mitigate natural hazards and disasters.

Estimates of precipitation serve as input to the hydrologic model used in this study. These estimates can be derived from many meteorological models and observational sensors such as rain gauges, satellites, weather radars, numerical weather forecasts, and combinations therein (i.e., multisensor estimates). They have traditionally been accepted as an accurate depiction of the initial state of the hydrologic system. The error characteristics of current precipitation estimates must be known; and moreover, the accuracy of these estimates must be improved before the uncertainty in hydrologic forecasts can be quantified and ultimately reduced. As pointed out in Droegemeier et al. (2000), hydrologic forecast uncertainty cannot be reasonably assessed until the uncertainty in the rainfall observations have been determined *a priori*. Entekahbi et al. (2002) identify the precipitation inputs as one of the major limitations to improved hydrologic predictability. Evaluation and quantification of uncertainty in model inputs is a major goal of this study.

Several quantitative precipitation algorithms are under development at the National Severe Storms Laboratory. It is highly desirable to learn which algorithms supply the most accurate inputs for hydrologic prediction purposes. This knowledge will guide future research in QPE algorithm developments. A true hydrologic evaluation can be accomplished most appropriately by testing

them directly; that is, by inputting them in the model and determining how well the model predictions of streamflow compare with observations. This methodology is attractive in that the experiment is being performed at the hydrologic scale of application. It can be argued, however, that model parameters can be judiciously adjusted or calibrated to account for errors in model inputs (e.g., QPEs). Systematic biases, originally present in the model inputs, can be mitigated or corrected in order to yield accurate streamflow forecasts. In short, adjustable model parameters can obscure uncertainties that were initially present in the model inputs. A new approach is proposed herein to account for uncertainty in the model parameters. Each input is "exposed" to the same degree of uncertainty in the parameters. An evaluation of model skill will elucidate the model input that results in the most accurate hydrologic predictions in light of the model parametric uncertainty. This concept is developed fully in Chapter III.

Input uncertainty is a component of the total prediction uncertainty in environmental modeling. It has been noted in several places, far too many to document herein, that forecasts of future states of the natural environment are plagued with inaccuracies. There are many reasons for this that depend on the physical system being modeled. In numerical weather prediction, for example, the future states of the atmosphere are very sensitive to accurate depictions of the initial conditions, leading to the theory of chaos (Lorenz 1963). A small error in the atmospheric initial state grows exponentially into much larger errors as

forecast times increase, drastically reducing the skill of the single, deterministic forecast (Lorenz 1969; Hamill and Colucci 1997). An alternative approach, called *ensemble forecasting*, first perturbs the initial conditions within bounds dictated by observational uncertainty. Each initial condition is used to produce a deterministic forecast, which is referred to as an ensemble *member*. An ensemble of forecasts is created by randomly selecting initial conditions that have an equal likelihood of representing the true state. These values are samples from the probability density function (pdf) that represents the distribution of the "true" initial state. Many simulations are then performed based on random selection of initial conditions from the pdf, where each member comprising the ensemble represents a single, deterministic possibility of the outcome. The group of members, or forecast ensemble, is treated collectively so that a probability of the forecast system state can be estimated. The spread of the forecasts represents the forecast uncertainty due to uncertain initial conditions while the mean of the forecasts is considered the best estimate (Stensrud et al. 2000). This probabilistic approach has shown improvements in the accuracy of weather forecasts at medium-range (Tracton and Kalnay 1993; Toth and Kalnay 1993) and short-range (Du et al. 1997; Hamill and Culucci 1997; Stensrud et al. 2000).

This kind of stochastic forecasting has attractive properties to risk analysts who often use cost-loss models based on forecasts with provided uncertainty bounds. As a simple example, consider a decision a rancher might need to make

in regard to transporting animals that may be in danger due to rising river levels. The cost of moving cattle out of a floodplain that doesn't end up flooding (i.e., a false alarm) may be insignificant compared to the penalty ensued in the event that the cattle aren't moved and are lost to a flood (i.e., a missed forecast). Future action may be taken if a probabilistic forecast of flooding is less than 50%. In this case, the penalty for a missed forecast outweighs the cost of needlessly transporting the cattle in a no-flood event.

Estimating the total uncertainty inherent to a forecast of an environmental variable involves the identification and quantification of uncertainty in all the following areas: model inputs, boundary and initial conditions, model parameters, model representations of physical processes, and observations of the system behavior. If successful, the pdf of future states of the system can be estimated accurately. An accurate pdf will greatly benefit decision-makers who are utilizing the forecasts for their application.

In numerical weather prediction, a majority of ensemble forecasts have been produced by members with different initial conditions. The underlying assumption is that the total prediction uncertainty is composed entirely of uncertainty in the initial conditions alone. No attempt is made in this scenario to quantify uncertainty in the model physics, physical parameterizations, numerics, or observations of the atmospheric state. Stensrud and Fritsch (1994) and Stensrud et al. (2000) include model uncertainty in addition to initial condition

uncertainty into separate ensembles. This combined approach has led to more dispersive forecasts, thus better representing the true pdf. Moreover, computational constraints have traditionally limited the number of simulations plausible for real-time applications. This has encouraged the creation of ensembles that produce the maximum amount of dispersion with the minimum number of simulations or members. The inclusion of additional components of uncertainty in environmental modeling, such as in the model physics, accomplishes this goal.

In hydrologic modeling, the total prediction uncertainty has been assumed to be largely, if not solely, a function of parametric uncertainty. This approach assumes that the model physics are perfect as well as the model inputs. Model parameter ensembles are created in an analogous fashion as initial condition ensembles described previously. In this case, parameter values are randomly chosen from the each parameter's pdf, or estimate thereof. Simulations are performed for each parameter realization to produce an ensemble of streamflow forecasts. This forecast ensemble is then used to derive the forecast distribution, measures of central tendency as well as the spread or uncertainty bounds of the forecasts. Little to no attention has been given to errors associated with model inputs (e.g., QPEs), model physics, or observations of streamflow. This observation is surprising given the wealth of studies in the literature documenting poor performance of radar-based rainfall algorithms. A major cornerstone of this

study involves the accounting for uncertainty in QPEs as an important component in hydrologic prediction uncertainty. Also, propagation characteristics of this input uncertainty are revealed.

This study is largely being undertaken in response to several recommendations promoted by CUAHSI (2002), NRC (2001) and USWRP PDT-9 (Droegemeier et al. 2000). The three basic research questions are stated and summarized below.

1. **How can comparisons of QPE algorithms be performed systematically and quantitatively?**

It is well acknowledged that the accuracy of streamflow predictions from a hydrologic model is heavily dependent on the accuracy of the rainfall inputs. Several efforts are underway (e.g., Gourley et al. 2002) to improve QPE by understanding the situations in which radar estimates can be erroneous and utilizing data from multiple sensors (e.g., infrared satellite, rain gauges, numerical weather output, and lightning flashes). Moreover, QPE from dual-polarization radars offers improvements where reflectivity-based rainfall estimates can be problematic due to hail contamination, ground clutter and anomalous propagation, and beam blockages. As these QPE algorithms are being formulated, it is vital to the developers to know the error characteristics associated with the estimates. Currently, there is no system in place that would systematically monitor the

progress of improvements in QPE algorithms in a statistically and hydrologically relevant way (Ciach and Krajewski 1999a,b; Ciach et al. 2000).

Traditionally, improvements in rainfall algorithms have been measured by comparing the remotely sensed QPEs to rain gauges at collocated grid points. This verification methodology lacks robustness for several reasons. First of all, true multisensor algorithms (e.g., Gourley et al. 2001) incorporate gauge data in their QPE schemes. Thus, it becomes a challenge to utilize truly independent data sources for verification purposes (Young et al. 2000). The use of these gauge values for verification purposes is thus inappropriate in this scenario. This situation could potentially be remedied if some gauges were withheld from the estimation scheme and used for verification instead. This evaluation design sacrifices the accuracy of the precipitation estimates and relies on a dense network of rain gauges.

Several references site the lack of accuracy with rain gauge point measurements (e.g., Zawadzki 1975). Underestimations are common with increasing wind speeds, and the gauges may only represent the rainfall that occurs in close proximity to the gauge. Highly variable rainfall may not be represented very accurately by a rain gauge network. Moreover, there are issues regarding the scales of measurement between rain gauges and remotely sensed rainfall. It has been noted that the sampling sizes between a typical radar pixel and a rain gauge orifice differ by about eight orders of magnitude (Droegemeier et al. 2000).

A methodology is proposed herein that provides the framework to evaluate QPE algorithms at the scale of application, a watershed. The probabilistic approach enables uncertainty estimates or confidence intervals to be assigned to the predicted hydrologic variables. A statistical analysis based on hydrologic predictions from differing QPE inputs as compared to observed streamflow reveals characteristics inherent to the algorithms. It is believed that this hydrologic analysis will serve as a tool for QPE algorithm assessment and refinement.

2. What are the predictability and limits-to-prediction in the hydrologic system?

The concepts of predictability are developed fully in NRC (2001). In general, if a forecasting system possesses predictability, then performance measures have been diagnosed and met when comparing the predicted variables to observations. One of the specific goals in predictability and limits-to-prediction outlined in the science initiatives in Hornberger et al. (2001) is to demonstrate the degree of predictability in hydrologic systems using probabilistic approaches to deal with uncertainty in modeling. This study accomplishes this goal.

An ensemble prediction system is implemented to ultimately quantify the total uncertainty in model predictions. Characteristics of prediction uncertainty for a given model structure are revealed through an examination of the relative

contributions of uncertainty from model inputs, model parameters, and combinations. The approach undertaken provides an accurate assessment of the forecast pdf, which is useful to end-users of hydrologic forecasts. It also reveals steps in the modeling process where uncertainties are large and can dominate the overall predictive uncertainty. This information is useful for focusing efforts on making improvements to the modeling process where they are needed most. Moreover, ensembles can now be constructed to minimize the number of simulations needed, and thus computation resources, while maximizing dispersion or spread. It is also shown how uncertainty in the model physics and observations of streamflow can be identified in this study. If uncertainties in the model inputs and parameters are accounted for, then any residuals must be due to uncertain model physics, imprecise observations, or numerical uncertainties.

3. How do uncertainties in model inputs propagate to errors in model predictions?

The final science question is posed in response to the second goal identified by the NRC (1999). The report states the need for the identification and understanding of the sources of uncertainty and the propagation of uncertainty in hydrologic systems. Presently, there is a lack of information regarding the error characteristics associated with radar QPEs. Moreover, many traditional hydrologic models rely on the principle of superposition to produce streamflow

using the unit hydrograph theory. These model structures thus dictate linear hydrologic responses to perturbations in rainfall inputs. The model used herein permits nonlinear hydrologic responses as this system behavior has been widely observed (e.g., Droegemeier et al. 2000). The study undertaken assesses the impact of uncertainties in radar-based QPEs on hydrologic predictions of streamflow. Propagation characteristics of input uncertainties are revealed by using historic hydrologic data sets that were collected over a natural hydrologic laboratory in Oklahoma.

An existing natural outdoor laboratory for hydrologic research is located in the Blue River Basin, near Blue, OK. (Fig. 1.1). The Blue River Basin drains about 1200 km^2. The headwaters of the basin are about 80 km away from the nearest weather radar while the basin outlet is over 200 km away from a radar. Radar-based rainfall estimates at far range may have errors due to beams overshooting shallow profiles of reflectivity, bright band contamination, and improper reflectivity-to-rainfall conversion equations. The long distance from radars makes this basin a good candidate for evaluating QPE algorithms at moderate to 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. In addition, the model used in this study has been carefully calibrated and utilized rather extensively on the Blue River Basin. The

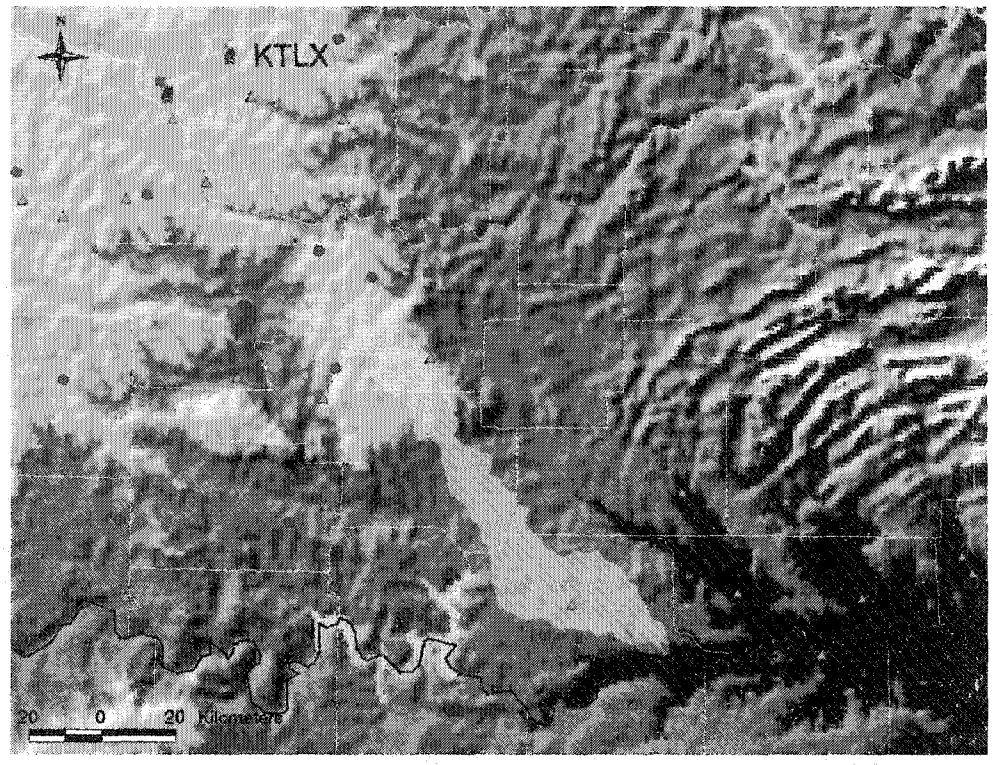


FIG. 1.1. The study area including the Blue River Basin (shaded in yellow), WSR-88D radars, Mesonet rain gauge locations (red filled circles), USGS gauging sites (light blue filled triangles), and a shaded relief map of Oklahoma.

model parameters have been refined manually using many observations of rainfall inputs and discharge combined with simulations of discharge. Nevertheless, as will be described in Chapter III, the methodology used herein offers a stochastic, automated approach for handling uncertainty contained in the parameter sets, and is thus readily applicable for use on basins with little or no prior modeling experience. The previous assembly of hydrologic datasets such as parameter maps needed in the model and observations of rainfall (by rain gauges and radar) and basin discharge greatly facilitate running the required simulations.

The focus in this natural outdoor laboratory is placed on the uncertainty in the rainfall inputs and to demonstrate a new methodology in comparing different QPE algorithms by examining the basin response in a probabilistic manner. The skill of forecast ensembles that are comprised of different inputs, model parameterizations, and combinations is assessed for several hydrologic events. The approach taken also reveals the optimum number of members needed to produce an ensemble that adequately describes the true forecast pdf. The methodology is applied here to the *Vflo* ™ hydrologic model with inputs provided by NSSL's Quantitative Precipitation Estimation and Segregation Using Mulitple Sensors (QPE SUMS) algorithm. The application of this methodology is directly transportable to other prediction systems that have inherent uncertainty, such as numerical weather prediction models or other QPE systems.

CHAPTER II. BACKGROUND

2.1 Quantitative Precipitation Estimation

2.1.1 Rain Gauges

Accurate measurement of precipitation in liquid and frozen phases is vital for commerce, agriculture, transportation, mitigating natural hazards, and environmental modeling. In agriculture, rainfall amounts combined with additional meteorological information often dictate the seasonal yield for a given crop. Monitoring snowfall accumulations and icing conditions impacts transportation on US highways and in airways. The costs associated with weather related delays are rather significant and must be considered in commerce (Stevenson 1997). The National Weather Service (NWS hereafter) is the agency in the US that is responsible for issuing warnings to the public prior to the onset of floods and flash floods. Presently, precipitation estimates from the Weather Surveillance Radar-1988 Doppler (WSR-88D) radar (Crum and Alberty 1993) are the primary tools used for monitoring impending flash floods. Other applications utilize precipitation estimates as initial and boundary conditions to model physical hydrological processes such as river discharge, chemical concentrations in lakes, rivers, and estuaries, the severity of storm surges from a hurricane, land surface and atmospheric interactions such as evapotranspiration, and prediction of future

atmospheric states including quantitative precipitation forecasting. The following section outlines traditional, current, and future methods of monitoring rainfall.

Conceptually, the simplest way to measure rainfall is to collect it in a device such as a rain gauge. Historically, the frequency of rain gauge measurements has been dependent on the observer manually measuring the depth of rainfall. The change in design of these instruments has largely been in response to the need to automate rainfall recordings and collect rainfall data in remote regions. Today, rain gauges can collect, record, and transmit rainfall amounts almost continuously through the use of small buckets that tip, drain, and activate a magnetic sensor, with larger buckets that use a weighing device, and also measurements of the total capacity in a small cylinder. Other types of non-collecting rain gauges estimate rainfall rates by measuring the scintillation in an optical beam produced by the raindrop shadows, acoustic waves caused by raindrops impacting a water surface, and the momentum/sound of raindrops falling on a styrofoam cone in a disdrometer (Nystuen et al. 1996). Typically, the diameter of a single rain gauge orifice is on the order of centimeters. The number of rain gauges that are deployed in a given region depends on the network configuration. The Oklahoma Mesonetwork (Mesonet), which is considered to be a very dense mesoscale weather observation network, operates 115 stations in the entire state (see Fig. 1.1). It can thus be said that the network monitors rainfall over approximately 8 m^2 of area in the entire state of Oklahoma, or 10^{-11}% of total

land area. Uncertainties in *domain-wide* rainfall measurement by rain gauge networks point largely to how well these "point" measurements characterize rainfall over the remaining ungauged area. In addition, a wealth of research has noted underreporting of rainfall amounts by gauges due to wind effects and loss of water mass during recordings.

Nystuen (1999) summarizes the error characteristics of different types of automatic rain gauges under differing weather conditions. In regard to collection-type gauges, they are typically placed on top of a surface as opposed to being dug into the ground. The gauge itself can cause turbulence, and irregular flow patterns result near and above the instrument. The perturbed wind flow often results in an undercatch that can be as large as 40% for high wind situations (see e.g., Wilson and Brandes 1979; Legates and DeLiberty 1993). Underestimates of heavy rainfall rates may also occur due to loss of rainfall in between measurements or "tips" as is the case with the tipping bucket rain gauge (Marselek 1981 and others). Underestimates of rainfall may also occur for low rainfall rates because a finite volume of water mass is needed before the small bucket can tip and trigger the sensor. The tipping bucket gauge thus has a 1-min rainfall rate resolution of ± 12 mm hr^{-1}. It is believed, however, that the largest errors with rain gauge measurements occur with *areal* estimates of precipitation (i.e., accounting for the rest of the land area that receives rainfall but isn't gauged as is the case with the Oklahoma Mesonet). The sampling error may be especially

prevalent for highly variable rainfall occurring over a sparse network of gauges. Better characterizations of spatially nonuniform rainfall may be possible using remote sensing platforms such as radar and satellite.

2.1.2 Radar

The advent of the WSR-88D radar has offered the potential to estimate rainfall at temporal resolutions equal to or higher than rain gauges and over very large regions at high spatial resolutions. The WSR-88D network employs approximately 120 radars that are jointly operated by the Department of Defense, Federal Aviation Administration, and National Weather Service. Each WSR-88D radar produces operational estimates of precipitation by measuring the power returned from hydrometeors illuminated by the radar beam. A radar reflectivity factor is computing by determining the ratio of power returned to the receiver to that transmitted by the antennae. Other factors entering the so-called radar equation quantify the constant operating characteristics of the radar and the range to the backscattering objects. The reflectivity data are collected at up to 14 elevation angles, along 360 radials (1° azimuthal resolution), out to a maximum range of 460 km (1 km range gate spacing) every 5, 6, or 10 minutes depending on the operating mode of the radar. The radar is said to be in precipitation mode when there is sufficient reflectivity observed by the radar. The occurrence of significant reflectivity underneath the radar umbrella initiates the precipitation

processing subsystem (PPS). This algorithm performs various quality control checks on the reflectivity data and ultimately generates a suite of precipitation accumulation products at each grid point out to 230 km in range. The algorithm remains very similar to its original design, which was developed at the NWS Hydrologic Research Laboratory in the 1980s (Fulton et al. 1998).

A multitude of adaptable parameters are used in the processing system with the intention of fine-tuning it to different geographic locations and seasonally variable weather conditions. Despite this flexibility, most radars utilize the same parameters to convert the radar-measured variable, reflectivity, to a rainfall rate. The process to do this reflectivity-to-rainfall conversion (Z-R) requires regressing an equation to observations of reflectivity and rain gauge accumulations. Most WSR-88D radars use the default, empirically-derived convective relationship from Woodley et al. (1975):

$$Z = 300R^{1.4}, \tag{2.1}$$

where Z is in $mm^6 mm^{-3}$ and R is in $mm\ hr^{-1}$.

Numerous studies have pointed out substantial variability in computed Z-R relationships (Wilson and Brandes 1979 and references therein). Multiplicative factors (e.g., 300 in (2.1)) and exponents (e.g., 1.4 in (2.1)) vary at different geographic locations, with different seasons and storm types, with rainfall

intensity, throughout different stages of storm lifecycles, and in the presence of vertical air density gradients (Dotzek and Beheng 2001). The physical reasoning for a majority of the reported variability lies in the sensitivity of the radar reflectivity factor to raindrop sizes and their distributions. The exact radar reflectivity factor (Z in $mm^6\ mm^{-3}$) can be determined only when the particle concentrations (n_i) and drop diameters (D_i in mm) within a sample volume are known (*Federal Meteorology Handbook No. 11* 1990):

$$Z = \sum_{i=1}^{\infty} n_i D_i^6 . \tag{2.2}$$

Theoretically, radar reflectivity can be determined exactly by measuring particle drop size distributions using a disdrometer. However, these in-situ measurements sample only a minute fraction of a precipitating cloud. WSR-88D sample volumes are many orders of magnitude greater than areas sampled by disdrometers. Thus, measurements of hydrometeor concentrations and their diameters within radar sampling bins cannot be known precisely. It may also be seen from (2.2) how sensitive the reflectivity factor is to larger drop diameters in a sample volume. For example, a drop 3 mm in diameter in a theoretical 1 mm^3 sample volume returns a reflectivity factor of 729 $mm^6\ mm^{-3}$. A drop 1 mm in diameter in the same unit volume only returns a reflectivity of 1 $mm^6\ mm^{-3}$ (2.1).

Despite having a reflectivity factor 729 times greater than the smaller drop, the bigger drop only has 27 times as much liquid volume. Variability in drop size distributions in rainfall results in uncertainty in Z-R equations (and thus quantitative rainfall rates) using linearly polarized radar data. Hydrometeor size spectra may be better resolved with dual polarization radar measurements (Seliga and Bringi 1976; Illingworth and Caylor 1989).

The accuracy of radar rainfall rates may also be limited due to a combination of the following: beam height increasing with range, sample volume increasing with range, unobserved precipitation growth below the beam, reflectivity rapidly decreasing with increasing height, and enhanced reflectivity below the 0°C isotherm. These sampling issues are a result of the radar transmitting and receiving power along paths of a constant elevation angle. Thus, the radar illuminates backscattering particles at higher heights relative to the ground with increasing range from the radar. The earth's curvature exacerbates height-dependent sampling at far ranges. This range-dependency wouldn't necessarily be problematic in radar QPE if the vertical profiles of reflectivity (VPR) in precipitating clouds were uniform. Unfortunately, this is rarely the case. In cool season, stratiform precipitation for example phase changes occur with height resulting in a highly nonuniform profile. The melting and subsequent enhanced aggregation of hydrometeors as they fall below the 0° C isotherm results in larger diameter particles and thus a drastic increase in reflectivity (2.2).

This zone of elevated reflectivity, or radar bright band (Austin and Bemis 1950), can cause precipitation overestimations up to a factor of 10 if it is not accounted for (Smith 1986). Above this layer, reflectivity structures of stratiform precipitating systems often reveal a rapid decrease of echo intensity with height. Due to the shallow nature of these precipitating clouds, radar samples of raindrops are possible only within a limited range near radar. Kitchen and Jackson (1993) found that radars typically underestimate surface rainfall by a factor of 10 at far range (e.g., 200 km).

Precipitation estimates from radar necessarily have biases that depend on range (Joss and Waldvogel 1990; Fabry et al. 1992; Kitchen and Jackson 1993; Smith et al. 1996; Seo et al. 2000). Several studies have adopted methods to account for the vertical variability of reflectivity by adjusting reflectivity measurements taken at some height above the ground so that they correspond to reflectivity (and thus rainfall rates) at the surface.

From observations, Joss and Waldvogel (1990) identified three VPRs representing different weather situations (thunderstorms, widespread rain, and drizzle or snow). They also computed percentages revealing how much the radar might overestimate or underestimate the melted water value measured at ground level. Later, Harrold and Kitchingham (1975) developed a procedure to compute a ratio of reflectivity measurements taken from two different elevation angles. This ratio is intended to adjust surface rainfall estimates where high-altitude

reflectivity samples were believed to be unrepresentative. Continuing the progress made by Harrold and Kitchingham (1975), Andrieu and Creutin (1995) proposed correcting for the range-dependence of radar precipitation estimates by utilizing a ratio of reflectivity factors taken from two different elevation angles at discrete ranges. A mean, singular VPR is then computed through an inverse solution. With this, a correction factor may be applied to the entire field of surface precipitation estimates based on the mean VPR and radar beam geometry. This method accounts for radar rainfall estimates decreasing with range due to beam overshooting and beam broadening. Joss and Lee (1995) incorporate real-time vertical reflectivity information to improve rainfall estimates over the Swiss Alps. Within 70 km of radar, reflectivity values from several heights are composited to determine a representative, mean VPR. They utilize this average VPR along with high-resolution terrain data and radar beam heights to extrapolate measured reflectivity values down to the surface. Beyond 70 km, they no longer assume the average VPR calculated near the radar represents reflectivity structures at far range. Thus, they incorporate climatological profiles to derive surface rainfall estimates in these regions. The climatological guidance is also applied when there is not enough precipitation near radar to compute a representative, mean VPR. The use of radar data alone for domain-wide, operational QPE may be insufficient in complex terrain or with shallow-structured precipitating clouds (Westrick et al. 1999; Gourley et al. 2002).

In the last two decades, many studies have demonstrated the capability of quantitatively estimating rainfall using multiple parameters that are available with polarization diverse radar measurements (e.g., Doviak and Zrnic 1993; Zrnic and Ryzhkov 1999; Straka et al. 2000; Bringi and Chandrasekar 2001). Dual-polarization radar has the ability to transmit and receive polarized waves in both the horizontal and vertical orientations. These measurements provide information on the size, shape, orientation, and thermodynamic phase of precipitation particles (e.g., Bringi and Chandrasekar 2001). As described above, single polarized radar estimates of rainfall are derived from horizontal reflectivity (Z_h). The polarization diversity enables the additional measures of differential reflectivity (Z_{dr}) and specific differential phase (K_{dp}), from which rainfall algorithms have been derived (Carey et al. 2000 among others). The differential reflectivity is the ratio of the amount of backscattered energy in the horizontal by a horizontally transmitted wave to the amount of backscattered energy in the vertical by a vertically transmitted wave. The specific differential phase is the difference between the amplitudes in forward scattering particles in the horizontal and vertical directions. Both measures are preferable to horizontally polarized reflectivity alone because they are less sensitive to variations in drop size distributions, and rainfall rates derived from these variables generally have better agreement with rain gauges (Ryzhkov et al. 1997, Carey et al. 2000 among many others).

It should be noted, however, that dual-polarization radar measurements offer improvements in QPE but cannot be considered perfect estimators of precipitation. Except under ideal atmospheric conditions, ground-based radars have a finite horizon similar to a line-of-sight limitation. Scattering is measured at greater heights with increasing range from the radar. The range-dependency is exacerbated by the following: wide spacing between adjacent radars, significant beam blockages near the radar by complex terrain, and the shallow nature of precipitating cloud structures common with cool season, stratiform rain, orographic rainfall and snowfall, and clouds that produce rainfall in the absence of frozen hydrometeors (i.e., warm rain processes). It is believed that no studies to this date have evaluated dual-polarization rainfall rates at ranges beyond 100 km from a radar. This kind of research will need to be conducted before rainfall algorithms based on dual-polarization variables can be considered for operational purposes. The following section describes how space-born sensors may be used to estimate rainfall by observing the radiative properties of the atmosphere and clouds from above.

2.1.3 Satellite

The deployment of operational weather satellites in the mid-1960s provided new opportunities for rainfall monitoring (Barrett and Martin 1981). Satellites are grouped by the nature of their orbit: 1) polar-orbiting and 2)

geostationary. The former environmental satellites typically follow low-level orbits between 500-1500 km above Earth's surface. They are sun-synchronous, so that they scan a new strip of the planet during each orbit. While their orbits allow them to investigate the entire globe, a given unit area on the surface is viewed, at most, twice a day. The geostationary satellites, often referred to as the Geostationary Operational Environmental Satellites (GOES), are placed at altitudes near 35,400 km above earth's surface. At this height, they follow a geosynchronous orbit that allows them to travel at the same rate as the rotation of the Earth. They appear to be fixed above a certain point above Earth's surface. The geostationary satellites monitor a limited field of view, but may do so frequently. Both types of satellites have been equipped with modern sensors capable of measuring radiation from earth. They measure energy emitted at wavelengths through atmospheric "window wavebands" (Barrett and Martin 1981). Most studies involved in rainfall monitoring have utilized data from radiation emitted at the visible (VIS), infrared (IR), and microwave (MW) channels.

Earlier developments in precipitation estimation by a remote sensing system aboard satellites have concentrated on tropical convective precipitation (e.g., Arkin 1979; Arkin and Meisner 1987; Griffith et al. 1981; Negri et al. 1984; Adler and Negri 1988). They were mainly concerned with identifying large-scale, long-term rainfall anomalies related to the general circulation of the atmosphere

(Rasmusson and Carpenter 1982). Many of these study regions are oceanic and are thus not equipped with a dense network of rain gauges or radars.

Adler et al. (1993) show how inclusion of MW data improved the performance of these techniques. In short, data from the Special Sensor Microwave Imager (SSM/I) were used to calibrate rainfall estimates from IR-based techniques. Despite poor temporal and spatial sampling of the polar-orbiting MW sensor, radiance in the MW portion of the spectrum responds to particles having sizes similar to precipitating hydrometeors. Brightness temperatures at 37 and 86 GHz provide a direct connection between measured radiance and raining areas. Other MW rainfall estimation techniques have produced encouraging results (Wilheit and Chang 1980; Spencer 1986; Olson 1989; Kummerow et al. 1989; Grody 1991; Liu and Curry 1992). However, since measurements are made only twice a day, MW data are of little use for techniques estimating precipitation at time resolutions on the order of several hours.

An automated technique for estimating rainfall called Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) (Hsu et al. 1997) uses satellite imagery and ground-based rainfall information. This technique is shown to work well for estimating large-scale precipitation patterns over radar-void oceanic regions. Like many other satellite-based studies, the main focus has been on precipitation associated with convection.

More recently, studies have refined spatial and temporal resolutions of estimation schemes using operational GOES satellites (e.g., Vicente et al. 1998). Satellite-based rainfall estimation has received much more attention recently due to the launch of the Tropical Rainfall Measuring Mission (TRMM) (Kummerow et al. 1998). Like GOES, the TRMM polar-orbiting satellite is capable of measuring radiances in the IR and VIS channels. In addition to passive measurement in the MW channels, the TRMM satellite is equipped with a precipitation radar that actively transmits and receives radiation in the MW channel. Data collected from this unique spectrum of observations has led to many published results too numerous to catalog herein. Future precipitation algorithms may include satellite data for operational applications in regions where other data sources (e.g., radar and rain gauge) may be inaccurate or unavailable.

2.1.4 Multisensor

A unique approach to operational QPE is to synergistically integrate data from all available precipitation sensors in a so-called multisensor algorithm. Previous efforts toward this goal have utilized rain gauge data to remove biases in radar-derived precipitation maps (see e.g., Wilson and Brandes 1979; Seo and Breidenbach 2002). Gauge-adjusted radar products are used operationally by the NWS for flood forecasting, verification of numerical weather forecasts of precipitation, and precipitation monitoring. The algorithm that performs most of

these functions has been developed at the NWS Office of Hydrology and is called the Multisensor Precipitation Estimator (MPE). The operational version of this algorithm performs similarly to the algorithm that was initially designed at the Arkansas Basin River Forecast Center (P1 technique) in that it uses radar reflectivity data adjusted locally by hourly rain gauge data.

Gourley et al. (2002) advances the multisensor concept with a study that examines stratiform precipitation in Arizona. The described study demonstrates how IR satellite data can be calibrated with radar data to produce domain-wide estimates of precipitation that are more accurate than using radar data alone. Today, this initial multisensor concept has been advanced as a real-time algorithm called Quantitative Precipitation Estimation and Segregation Using Multiple Sensors (QPE SUMS). Precipitation estimates from this algorithm will also be evaluated herein, thus a brief algorithm overview follows.

The first procedure employed in QPE SUMS is quality control of the input radar reflectivity data. Observations of raw reflectivity data indicate several instances of backscatter from non-meteorological targets such as trees, terrain features, buildings, towers, vehicles, etc. This kind of "contamination" can carry over into precipitation products and yield erroneous accumulations. QPE SUMS checks isolated targets for meteorological coherence. Specifically, the clutter removal algorithm identifies and removes high reflectivity values that have no vertical structure. These suspect grid points must also have zero velocity

signatures because stationary targets are more common with non-weather phenomena.

After the input data from all available radars have been quality controlled, they are mosaicked on a common grid using the reflectivity that was measured closest to the ground. The algorithm then determines the phase and character of precipitation at each grid point. This segregation procedure determines which sensors will be activated for QPE. Much research has been reviewed and performed on the accuracy of each precipitation sensor for different weather scenarios. For example, as described in section 2.1.2 radar reflectivity data can be unrealistically high due to hydrometeors that are undergoing phase changes. Frozen hydrometeors melt, become water-coated, and aggregate which produces high reflectivity returns. QPE SUMS is designed to automatically search for this melting layer, and then discard the contaminated reflectivity values that were measured in this region. The bright band identification algorithm is described in detail in Gourley and Calvert (2003). Moreover, the heights at which the melting layer was found from the radar observations are used to map out the rain-snow line by comparing the melting layer heights with terrain heights. This enables the algorithm to determine the precipitation phase at the surface. This component alone is important for transportation purposes (e.g., snow removal operations) and characterizing the land surface infiltration rates, roughness, and runoff potential in hydrologic modeling.

Radar reflectivity data are also known to have range-dependent biases especially with shallow, stratiform precipitation. First of all, a convective-stratiform partitioning algorithm is applied to the data to determine which regions are receiving rainfall from stratiform versus convective clouds. In short, this subroutine identifies convective grid cells as those that have reflectivity values greater than 60 dBZ or reflectivity greater than 35 dBZ at temperatures colder than -10° C. The latter requirement is tied to microphysics as high reflectivity displaced up to cold temperatures indicates the presence of significant updraft velocities common with convection. The remaining non-convective grid cells are deemed as being stratiform. QPE SUMS treats the stratiform echo differently in the "radar-only" module than in the "multisensor" module. The radar-only routine will use a different *Z-R* equation that is more applicable for stratiform precipitation (Marshall and Palmer 1948). The multisensor algorithm builds upon the IR satellite and radar regression technique that was prototyped in Gourley et al. (2002) to supply meaningful precipitation rates to stratiform pixels at intermediate and far range where radar estimates of rainfall are known to be inaccurate. That is the primary difference between the two different precipitation outputs in the QPE SUMS algorithm.

Next, QPE SUMS incorporates rain gauge observations on an hourly basis from the Oklahoma Mesonetwork (Mesonet) to produce gauge-only, mean field bias adjusted rainfall, and local bias adjusted products. The gauge-only analysis

incorporates all available rain gauge observations in a domain. Initially, a simple quality control (QC) step is applied to the data. The QC routine is capable of checking the magnitude, temporal consistency, and spatial consistency of each rain gauge report. Currently, only the magnitude check is applied to the data such that gauge accumulations are ignored if the accumulations exceed 203 mm/hr. If the reporting interval of the gauges is more frequent than one hour, then the accumulations are aggregated to produce hourly accumulations. The point estimates are analyzed on the 1x1 km QPE SUMS common grid using a Barnes objective analysis scheme. The two parameters optimized in the gridding process are the weighting function and the radius of influence. Subjective analysis is used for observed precipitation events to determine the best parameter settings.

The multisensor and radar-only products are both adjusted using a mean field bias correction (–G hereafter) based on the comparison of the nearest grid point values to the gauge locations. This adjustment technique is intended to remove domain-wide biases on an hourly basis that may be due to improper Z-R equations, overestimation from hail contamination, or underestimation from virga. The mean field bias is computed as follows:

$$\beta = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{G_i}, \qquad (2.3)$$

where β is the bias factor, R_i is the hourly multisensor or radar-only accumulation, G_i is the hourly gauge accumulation and N is the number of gauges reporting for that hour. Some checks are placed on the bias value to ensure that no unreasonable biases are computed. If the mean of the gauge accumulations (G_i) over the entire domain is zero and the mean of the estimator (R_i) at the gauge locations is zero, then the bias factor is set to a value of one. If the mean of the gauge accumulations (G_i) is zero and the mean of the estimator (R_i) at the gauge locations is nonzero, then the bias factor is set to a value of zero. Otherwise, the entire grid of radar-only and multisensor hourly rainfall is divided by the bias factor to produce the radar–mean field bias adjusted and multisensor–mean field bias adjusted products. Like all other QPE products, these are aggregated to produce long-term accumulations.

The multisensor and radar-only products are adjusted on an hourly basis using a spatially nonuniform bias adjustment technique (-LG hereafter). This adjustment is intended to remove nonuniform biases that may be due to improper *Z-R* relationships, range-dependency in QPEs from reflectivity profiles that decrease with height, and contamination from hail, birds, ground clutter, chaff, and other echoes from nonweather targets. First, the difference between the gauges and the estimators is computed at each gauge location (e.g., *G-R*). A local bias field is then computed for the 1x1 km QPE SUMS common grid using the point (*G-R*) values. The parameters used in the Barnes objective analysis scheme

discussed above are used here to analyze the point (*G-R*) values at the gauge locations to the entire grid. The weighting parameter and radius of influence are dependent on the gauge spacing with a given gauge network. Finally, the local bias field is added to the multisensor and radar-only hourly products to yield the radar-local bias adjusted and multisensor-local bias adjusted QPE products.

A third gauge adjustment strategy initially removes the mean field bias to account for domain-wide inaccuracies mainly due to inappropriate *Z-R* equations being applied. After the bias is removed, the local bias adjustment as described above is implemented to both radar-only and multisensor products. These products are referred to as local bias adjusted QPEs with bias removed (-LGG hereafter).

The suite of QPE SUMS products includes gauge-only (GAG), multisensor (MS) and radar-only (RAD) precipitation with the option of all gauge adjustment techniques (-G, -LG, -LGG) applied to each algorithm totaling nine different outputs. Each of these algorithms will be input to the hydrologic model in order to determine their uncertainty characteristics from the hydrologic perspective.

2.2 Hydrologic Models

The definition of a model may be made in several different contexts. Here, an environmental model is defined as the process of applying preconceived

ideas on data to make a prediction in which an informed decision may be made. The basic components of a model include inputs, parameters, state variables, and observations of the state or output. Models may be classified according to the following distinctions: 1) the treatment of the model inputs and parameters as a function of space and time, 2) the extent of physical principles that are applied in the model structure, and 3) the presentation of results in either a deterministic or stochastic manner.

In regard to the first criterion, hydrologic models have historically treated basin parameters (e.g., soil infiltration rates) and inputs (e.g., precipitation) as a single, basin-averaged number. These *lumped* models do not account for the spatial variability of model parameters or inputs within a basin. More recently, data sets related to basin hydrologic properties and precipitation inputs have been made available at high spatial resolutions through advances in remote sensing. *Distributed* models utilize these maps and thus account for the intra-basin variability of parameters and inputs. A hybrid between these two approaches is to apply a model at the sub-basin level. *Semi-distributed* models (e.g., Boyles et al. 2001) thus account for differing degrees of the spatial distribution of inputs and parameters.

Beck (1987) provides a complete overview on the different classes of models. The model classes are differentiated based on the concept of *identifiability*. Identifiability is the ability to determine unambiguously if a

hypothesis in the model structure is false (Young 1978). A model developer, for example, may want to test several hypotheses by changing the interaction between parameters and state variables, i.e., the governing equations. In order to determine the success or failure of a given hypothesis, the modeler would then examine the prediction from the new structure using the same initial/boundary conditions used in the previous structure. In an over-parameterized model, the new prediction may be very similar to the former model forecast. This lack of sensitivity to a different model structure may indicate that the model has surplus content, i.e., the modeler is unable to effectively falsify the tested hypothesis. The definition of this characteristic in modeling is called a lack of identifiability and is useful in defining the following three classes of models.

The Class I model describes the system behavior in as much detail as conceivable. The model structure (composed of state variables and parameters) contains all known physical laws that are thought to be important to the dynamics of the system. This model has been referred to as a physics-based, or a whitebox model. The whitebox analogy refers to the modeler's ability to see directly into the model structure. Given a set of input data, the output is a direct solution of the physical principles that were applied to the data in the model. Often these logical steps or partial differential equations can be quite numerous requiring lots of parameterizations. Over-parameterization may result and lead to ambiguities when testing new physical hypotheses. The Class I model is prone to a lack of

identifiability and thus greater structural uncertainty. On the other hand, the physics-based model has flexibility in that it can predict future states of behavior that have not been observed.

The Class III model is based entirely on past system behavior, i.e., observations of input and output, and employs no physical principles. Instead, functions are derived that directly relate the input into a system to the output. This model has been referred to as a statistical or blackbox model. The blackbox analogy suggests that the model structure doesn't reveal the reasoning behind the system response to a given input. Predictions from the Class III model are limited by the memory of observations of previous system behavior. This model is incapable of predicting a future that has not been observed. Moreover, there may not be enough historical information about the system being studied in order to form the model basis (e.g., basins with no previous records of discharge). This leads to the following model of compromise.

The Class II model is the middle ground between the Class I and III models. In this case, a set of concepts is still applied; however, the logic is calibrated based on the observed input and output data. This model is referred to as a conceptual or graybox model. Whitebox suggests that the modeler can observe the underlying physics within the model structure as in the Class I model. These hypotheses are altered based on observations of the relationships between input and output. Often, the modeler may not fully understand why these

relationships do not agree with predictions based on physical principles alone. Inputs, parameters or the model structure itself are adjusted so that the predictions match the observed system behavior more closely. The inability to model the system perfectly without any calibration obscures the model structure and colors it gray. Similar to the Class I model, the Class II model is capable of predicting unobserved future states. However, the number of concepts has been reduced. Thus, the model is said to be more identifiable. This graybox model has less structural uncertainty than the Class I model and more flexibility than the Class III model, and thus makes it preferable for environmental modeling.

The final distinguishing characteristic between models is the method in which the results are presented. Deterministic models yield a unique output vector for a given set of initial/boundary conditions. This kind of output may in fact be useful if the inputs, parameters, and the physical system being modeled are known precisely. In environmental modeling, this kind of certainty is rarely the case. Typically, there is some uncertainty associated with the inputs, parameter values, model structure, and observations of output. This uncertainty can be included in the predicted variables and presented in a stochastic or probabilistic framework. This provides for the range of possibilities and the likelihood that a given prediction will occur. Many references have shown how probabilistic information is useful in decision-making, especially when distinguishing between

accuracy and value in a cost-loss ratio situation (see e.g., Murphy and Ehrendorfer 1987).

There are many hydrologic models, much too numerous to catalogue herein, that may be classified according to the definitions provided above. The interested reader is referred to the following list that summarizes the rainfall-runoff models that are used most predominantly for research and operations: CASC2D (Julien and Saghafian 1991; Ogden and Julien 1994; Julien et al. 1995), Systeme Hydrologique European (SHE) (Abbott et al. 1986a; b), Distributed Hydrology Soil Vegetation Model (DHSVM) (Wigmosta et al. 1994), Precipitation-Runoff Modeling System (PRMS) (Leavesley et al. 1983), Sacramento Soil Moisture Accounting Model (SAC-SMA) (Burnash et al. 1973), and HEC–HMS model (Hydrologic Engineering Center 2000). The hydrologic model used in this study is a commercial model developed at Vieux and Associates, called *Vflo* ™ (Vieux and Vieux 2002).

In reference to the previous classifications of environmental models, the *Vflo* ™ model treats parameters and inputs in a spatially distributed fashion, and flow simulations are allowed to vary with time, i.e., unsteady. The model is considered a Class II model because the governing equations are derived from conservation principles while the model parameters are calibrated by scalars using the Ordered Physics-based Parameter Adjustment method (OPPA) described by Vieux and Moreda (2003). Recall that the Class II model is flexible in that it can

predict future states that have not been observed while the model is identifiable and thus has minimized model structural uncertainties.

Prior to performing simulations, distributed parameter maps that describe or are at least analogous to the following soil properties must be obtained for the basin of interest: saturated hydraulic conductivity, initial degree of soil saturation, and soil suction at wetting front. These soil properties may be inferred from State Soil Geographic (STATSGO) survey maps that are freely available for most basins in the US in a geographical information system (GIS) format. The aforementioned soil properties are required for proper simulation of soil infiltration rates that are then used in the formulation and solution of the conservation equations (described below). A digital elevation model (DEM) is needed for deriving the slope map, watershed boundaries including the basin outlet, flow direction map, and the location of DEM-defined streams. Most GIS software is capable of performing the aforementioned operations. The final parameter map that must be obtained is a spatially distributed depiction of the Manning roughness coefficient. This hydrologic parameter may be inferred from readily available land use/land cover maps that are typically derived from aerial photography or, more recently, from high-resolution radiometer measurements aboard polar-orbiting satellites. Typically, these maps have coded values of land use/land cover. Published tables in Chow et al. (1988) are used to convert the coded values to actual Manning roughness values. Finally, channel characteristics

such as the cross-sectional area, geometry, channel side slope, and hydraulic roughness must be specified for each stream.

The 1-D conservation of mass (2.4) and momentum equations (2.5), commonly referred to as the Saint-Venant equations, are used to derive the governing equations in the *Vflo* ™ model.

$$\frac{\partial h}{\partial t} + \frac{\partial (uh)}{\partial x} = r - i, \tag{2.4}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + g(\frac{\partial y}{\partial x} - S_o + S_f) = \frac{u(r-i)}{h}, \tag{2.5}$$

where u is the 1-D component of velocity, h is the flow depth, r the rainfall rate, i the soil infiltration rate, g the acceleration due to gravity, S_o the bed slope, and S_f the friction slope. The momentum equation (2.5) is simplified by making the assumptions utilized in the kinematic wave analogy. The local acceleration, horizontal momentum advection, hydrostatic pressure, and forcing terms on the right hand side of (2.5) are all assumed to be at least an order of magnitude smaller than the friction and bed slope terms. Thus, the momentum equation (2.5) reduces to:

$$S_o = S_f. \tag{2.6}$$

Neglecting the aforementioned terms to form the kinematic wave equations assumes that the slope of the energy grade line and water surface elevation are parallel to the land surface slope. It is also assumed that the flow is uniform within the element being considered. In nature, water flow over flat terrain may be subcritical. In this event, changes in water levels may result in waves propagating upstream. These backwater effects can only be accounted for in the local acceleration, advection of horizontal momentum, and hydrostatic pressure terms. All of these terms are neglected in the kinematic wave model, thus solutions of flow depth in areas where backwater effects may be important are not sufficient. Next, an appropriate relationship must be utilized that relates the flow velocity to the flow depth such as Manning's equation in SI units:

$$u = \frac{R^{2/3} S_f^{1/2}}{n}, \tag{2.7}$$

where R is the hydraulic radius (defined in (2.8) below) and n is the Manning roughness coefficient.

$$R = \frac{hw}{2h + w}, \tag{2.8}$$

h is the flow depth and w is the elemental flow width. The hydraulic radius in a rectangular channel can be approximated by the flow depth when the flow width is assumed to be much greater than the flow depth. Combining this approximation with the result from (2.6) reduces Manning's equation to:

$$u = \frac{h^{2/3} S_o^{1/2}}{n}. \tag{2.9}$$

The simplified form of the momentum equation above can be substituted into the continuity equation (2.4) and rearranged to yield the following governing equation used for overland flow in the *Vflo* ™ model:

$$\frac{\partial h}{\partial t} + \frac{S_o^{1/2}}{n} \frac{\partial h^{5/3}}{\partial x} = r - i. \tag{2.10}$$

The continuity equation may also be expressed in terms of the cross-sectional area A instead of the flow depth h. This leads to the formulation of the conservation of mass for channelized flow:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q, \tag{2.11}$$

where Q is the channel flow rate and q is the rate of lateral flow entering the channel per unit length. Analogous to the treatment of (2.4), it can be shown that substitutions and approximations of the momentum equation effectively relate the flow rate Q to a cross-sectional area A using the Manning equation.

The forcing function on the right hand side of (2.10) is the difference between the soil infiltration rate and rainfall rate, or rainfall excess. The Green and Ampt equation is used in the treatment of soil infiltration in the *Vflo* ™ model. The infiltration rates are equal to rainfall intensities as long as the rainfall intensities are less than the potential infiltration rates. When the rainfall rate equals and exceeds the potential infiltration rate, the soil is saturated and water becomes ponded at the surface. This ponded water is now available for overland or channelized flow to the adjacent, downstream grid cells. The potential infiltration rate is given as follows:

$$i(t) = K\left(\frac{\psi\Delta\theta}{I(t)} + 1\right), \tag{2.12}$$

where K is the saturated hydraulic conductivity, θ is the soil moisture content, ψ is the soil suction at wetting front, and $I(t)$ is the cumulative infiltration. The cumulative infiltration prior to soil saturation is calculated from the rainfall rates accumulated over time. When the rainfall intensity exceeds the infiltration rate in

(2.12), the time to ponding (t_p) and the associated cumulative infiltration at ponding time (I_p) have been reached and are used below. After ponding has occurred, the following equation describes the cumulative infiltration I:

$$I - I_p - \psi\Delta\theta \ln\left(\frac{\psi\Delta\theta + I}{\psi\Delta\theta + I_p}\right) = K(t - t_p)\,. \qquad (2.13)$$

I must be solved for implicitly using a method such as a Newton iteration. At this point, the cumulative infiltration may be inserted in (2.12) to yield the infiltration rate after ponding has occurred. This completes the formulation of the governing equations (2.10-2.11) for overland and channel flow in the *Vflo* ™ model. The next section describes the numerical procedure for obtaining solutions to the partial differential equations above.

The governing equations describe the behavior of water depths in channel and over land using continuous differential operators. Solving these equations requires either an exact, analytic solution or an approximation using finite numerical methods. One technique used in environmental modeling is to approximate the derivates in a partial differential equation using the finite difference method. Another approach that was developed initially for applications in mechanical engineering is the finite element method. The interested reader is referred to Segerlind (1984) for a detailed treatment of this

method. Instead of approximating the differential operators as is done in the finite difference method, actual solutions to the governing equations are approximated while maintaining the original differential operator. The method of weighted residuals discretizes the function space by using a finite number of points or nodes interconnected by elements. The solution method introduces a function with well-known characteristics such as piecewise Lagrange polynomials. These basis functions are then multiplied by coefficients at nodal locations and summed over the entire domain. The method of weighted residuals then seeks to minimize the residual, or the amount to which the trial solution fails to satisfy the original governing equation by forcing it to zero when considering the weighted average of the residual over the entire domain. The residual at nodal points will not be identical to zero, while the summation of the residual over the domain should approach zero. The Galerkin formulation of the finite element method uses the basis functions to act as weighting functions when driving the residual to zero. This procedure provides for the approximation of the flow depth over land and in channels at each node for a particular time step.

The solutions to the governing equations are marched forward in time using a finite difference approximation called the weighted Euler approach generalized below:

$$U^{n+1} = U^n + (\Delta t)\left[\theta F^{n+1} + (1-\theta)F^n\right], \qquad (2.14)$$

where U represents the dependent variable at some time ($t = n, n+1$, etc), F is a function describing the relationship between the dependent variable and time, t. The value of θ determines the weighting placed on the value of the dependent variable at time n versus $n+1$. When θ is set to 0, the forward Euler method results, and U^{n+1} may be solved for explicitly. On the other hand, a backward Euler method is obtained by setting θ equal to 1. In this case, U^{n+1} must be solved for implicitly using a method such as Newton iterations. The forward and backward methods are first order accurate, while setting the θ parameter to 0.5 results in second order accuracy. The backward Euler method produces stable results regardless of the size of the time step, thus it is said to be unconditionally stable. The forward Euler method, on the other hand, may become unstable with large values of Δt, thus it is said to be conditionally stable. The *Vflo* ™ model employs the forward scheme due to its computational efficiency. The stability of the solutions is maintained by constraining the length of the time step by the Courant condition. The Courant number is defined as follows:

$$\nu \equiv \frac{c\Delta t}{\Delta x}, \tag{2.15}$$

where ν is the dimensionless Courant number and c is the velocity at which a gravity wave propagates across the smallest element, Δx. The dynamic wave celerity is given as:

$$c = \sqrt{gh}\,, \tag{2.16}$$

where g is the acceleration due to gravity and h is the depth of the water. For stability, the Courant number ν must be less than 1. An appropriate value for Δt is found each iteration using (2.15-2.16). This adaptive time stepping technique ensures the solution that is propagated forward in time has stable characteristics.

2.3 Model Calibration

Uncertainty in environmental modeling arises due to incomplete knowledge about a physical process that relates the input to the output, inaccurate measurements of the input and output, and imperfect model parameterizations. Model calibration in this context is the process of reducing the ranges in the parameter sets so that the parameter uncertainty is reduced. Model calibration strategies are quite numerous and variable, but can generally be classified as being either manual or automatic. Manual techniques require a human to interactively run simulations using different values for a given parameter. The human then makes a decision as to the parameter value that allows the model to

simulate the system behavior most accurately. Parameter settings that yield solutions that are classified as unrepresentative of the observed system behavior are termed nonbehavioral and are rejected. This iterative process can be inefficient and even ambiguous if there is more than one minimum present in the parameter space. Automatic methods, on the other hand, vary in complexity but generally involve performing numerous solutions using different parameter settings. Past observations of rainfall and runoff are utilized to determine the accuracy of each prediction for a given model and its parameter values. The decision on the "goodness of fit" in this case is based on some performance criteria such as minimizing an objective function, or several objective functions as in multicriteria schemes (Gupta et al. 1998; Yapo et al. 1998) that are defined by the modeler.

Much attention in the hydrologic literature has focused on the successive adjustment of model parameters to observations of rainfall and runoff data through automatic methods (see e.g., Sorooshian and Dracup 1980; Beven and Binley 1992; Duan et al. 1992 among many others). These automatic methods can be further subdivided by their choice of a single, globally optimized parameter vector, or a *set* of parameter vectors that model the system adequately within a realistic range of observations. Optimization involves defining an objective function or functions, designing an algorithm to search the parameter space, utilizing historic records of rainfall and runoff for calibration, and setting

convergence criteria to terminate the searching procedure. Additional subdivisions can be made in automatic calibration methods by the exact searching procedure that is implemented.

In general, the searching procedures are designed to minimize (or maximize as appropriate) an objective function by identifying a minimum (maximum) in the parameter response surfaces. Some examples of these algorithms include the pattern search method (Hooke and Jeeves 1961), the Rosenbrock method (Rosenbrock 1960), the simplex method (Nelder and Mead 1965), the random search method (Karnopp 1963), and the adaptive random search (ARS) (Masri et al. 1978). The shuffled complex evolution method developed at the University of Arizona (SCE-UA) (Duan et al. 1994) is based on a synthesis of the best features from several existing procedures such as the ARS method (Pronzato et al. 1984) and the genetic algorithm reported in Holland (1975) plus complex shuffling (Duan et al. 1992; 1994). The Duan et al. (1992) study reports on the following common problems encountered in automated model calibration when a global optimum is desired: 1) more than one minimum (both local and global) in the parameter response surface, 2) "roughness" in the response surface due to discontinuous derivatives, 3) poor sensitivity of the response surface caused by insensitive parameters and interactions with other parameters, and 4) poorly defined minima due to a response surface with valleys and plateaus. In addition, difficulties may arise when the residuals are

autocorrelated and heteroscedastic (Sorooshian and Dracup 1980). It is shown through case examples that the SCE-UA algorithm does indeed locate the global minimum in the response surface as desired. However, such automated methods will have limited success for Class II models that are more physically-based, have more parameters, and will have intercorrelated parameters because they are related to physical principles such as conservation of mass and momentum.

A different paradigm in automated calibration exists that is based on the premise that environmental modeling inherently includes errors in the observations and measurements from which calibration is based, in addition to errors in our physical understanding of a process, i.e., errors are present in the model structure. Given this uncertainty, there is little reason to believe that the algorithmic procedures described in the preceding section will identify the one set of parameter values that represents the "true" parameter set. Probabilistic calibration methods recognize that there may be several different regions in the parameter space that result in equally acceptable simulations of streamflow for a catchment.

The generalized likelihood uncertainty estimation (GLUE) method (Beven and Binley 1992) assesses the *likelihood* that a given parameter set adequately simulates the observed system behavior. This concept of "equifinality" (Freer et al. 1996) gives rise to their concept of computing the probability that a given parameter set will yield an acceptable result. The GLUE method ultimately

computes a likelihood distribution for the predicted variable. This procedure initially requires the assignment of ranges for each parameter. An ensemble is created by sampling the parameter space using a methodology such as Monte Carlo simulation. For each parameter set, predicted streamflow is compared to observations and given an objective assignment of its "goodness of fit". As is the case with all objective functions, the choice of the likelihood measure is a subjective one. Van Straten and Keesman (1991) provide a summary of performance criteria that are often used. In model calibration, criteria may be established for each performance measure to assess whether a given model (including the parameter set) yields acceptable behavior or not. This set-membership approach was first described by Keesman (1990). After each parameter set has been classified as either "behavioral" or "nonbehavioral", the behavioral models are weighted and rescaled to give a cumulative sum of 1.0. The weighting concept is introduced in the event that some information about the likelihood distributions becomes available after the initial calibration step and there exists a need to sample the posterior distribution. The weighting may be established based on the importance of prior likelihoods (see Tanner 1992) or by using a random walk that adapts to the true probability distribution as in the Metropolis algorithm (Metropolis et al. 1953; Kuczera and Parent 1998). Regardless of the weighting scheme applied, the likelihood measures for each model are scaled such that the sum equals unity. The requirements for a discrete

probability distribution have been met using this procedure; thus providing for the computation of uncertainty bounds or quantiles (e.g., 90% simulation limits) of predicted streamflow. In essence, the GLUE methodology provides the framework to: 1) perform model calibration, 2) assess prediction uncertainty due to parameter uncertainty for a given model structure by casting predictions in a probabilistic sense, and 3) evaluate the sensitivities of predictions to each parameter as in Spear and Hornberger (1980). The methodology in Chapter III describes how this probabilistic procedure is expanded to include the uncertainties in the rainfall inputs for assessment of the total prediction uncertainty. It will also be shown how this ensemble approach can identify model structural errors.

2.4 QPE Evaluation Methodologies

2.4.1 Point Measurements

Rainfall rate algorithms from radar have been evaluated with rain gauges and disdrometers at time scales on the order of minutes through simulations and observations (Bolen et al. 1998). The techniques used in the comparisons vary in complexity from bias, average difference and relative dispersion as in Wilson and Brandes (1979) up to making considerations for the space-time differences between radar measurements and surface gauge observations. At an instant, a radar measurement of precipitation is made at some height, z, above the ground. There is a finite amount of time in which the radar-observed precipitation falls to

the surface and is thus observed by a rain gauge or disdrometer. In addition, horizontal winds may advect the raindrops so that they are displaced in space from the original measurement location. Approaches have been devised to shift the radar observations in time and search for an optimal area of radar data that likely contributed to the drop size distributions (DSD) measured at the surface (Bolen et al. 1998). The described adjustments seek near equivalence between variables measured by radar and by ground-based rain gauges or disdrometers. Parameters in the equations that relate the measured variable (e.g., Z_h, Z_{dr}, K_{dp}) to a physical process (rainfall) are often adjusted or fine-tuned in accordance with the disdrometer measurements (Schuur et al. 2001 among others). It can thus be noted that many rainfall rate algorithms have been designed and carefully calibrated to match disdrometer or rain gauge measurements as closely as possible. Comparisons between radar rainfall rates over a volume and gauge catchments in a small orifice are complicated by vertical variations in DSD owing to precipitate phase changes or precipitation growth below the radar beam, horizontal wind drift, underreporting of rainfall rates by gauges due to instrument-induced turbulence and splash-out (see previous discussion on rain gauges).

2.4.2 *Hydrologic Simulations*

Evaluation of the hydrologic response to differing rainfall algorithm inputs (observed and predicted) for flash flood events has recently been explored (Peters

and Easton 1996; Frank et al. 1999; Ogden et al. 2000; Warner et al. 2000a; b; Yates et al. 2000; Sanchez-Diezma et al. 2001; Westrick and Mass 2001). The modeling system approaches to evaluation often recognize the nonlinearity in the rainfall-runoff transformation (Droegemeier et al. 2000); and moreover, they note the complexity of the propagation of seemingly small errors in rainfall fields to much larger errors in predicted discharges (Frank et al. 1999; Ogden et al. 2000). The potential drawback in using such an approach is the introduction of model uncertainty, which is present in physical representations of most environmental systems. The nonlinear nature of the physical transformation of rainfall to runoff may be exacerbated by uncertainties in the modeling process itself, further complicating the evaluation capabilities of a hydrologic prediction system.

QPE verification using hydrologic model simulations is also complicated by the dependence of the hydrologic response on calibrated parameters. Yates et al. (2000) compares flood discharge simulations for a flash flood event that occurred on Buffalo Creek outlet on 12 July 1996. The PRMS model (refer to Section 2.2) simulated the basin discharge using inputs based on WSR-88D radar reflectivity, Z_h and K_{dp} measurements from the National Center for Atmospheric Research's (NCAR) S-Pol radar, and quantitative precipitation forecasts from a dynamic model and an automated algorithmic system. The study manually calibrates the saturated hydraulic conductivity parameters for each sub-basin using the S-Pol K_{dp} rainfall inputs. These inputs were used for calibration because

they are assumed to be the most accurate. Not surprisingly, the K_{dp}-based QPE algorithm matched the peak estimate of discharge for the flash flood event.

The previous example shows how model results can be dependent on the calibration data set from which they were derived, which are potentially misleading. Additional calibration procedures for different rainfall algorithms are rather cumbersome and have been largely ignored in studies evaluating different rainfall algorithms with a hydrologic model. It is suggested in Freer et al. (1996) that the GLUE technique be expanded to include the uncertainties associated with different rainfall inputs. Extension of the GLUE methodology to account for differing rainfall inputs provides a consistent methodology to evaluate the hydrologic response to each input *independently*. Through extension of the GLUE methodology used on the deterministic *Vflo* ™ model to generate likelihood distributions, the following chapter describes a new methodology that can determine which QPE algorithm is more accurate from a hydrologic perspective and at the scale of the integrating watershed. At this time, no known studies have evaluated QPE algorithms using a hydrologic model in a probabilistic way that is independent of the calibration data set. This ensemble approach is also instrumental in identifying model structural errors.

CHAPTER III. METHODOLOGY

3.1. Research Objectives

The purpose of this study is to answer the following three science questions that were posed in Chapter I: 1) How can comparisons of QPE algorithms be performed systematically and quantitatively? 2) What are the predictability and limits-to-prediction in the hydrologic system?, and 3) How do uncertainties in model inputs propagate to errors in model predictions? An ensemble approach to hydrologic modeling is developed to include the uncertainties in the input rainfall estimates and model parameters. The combined effect of these uncertainties provides a better estimate of the total prediction uncertainty. Model parameter ensembles are produced for each quantitative precipitation estimation (QPE) algorithm. The combined results from these ensembles are compared to each other and to observations in order to evaluate the accuracy of the inputs at the hydrologic scale. This evaluation methodology is believed to be unique and can apply readily to other modeling applications.

Chapters I-II demonstrate a need in hydrometeorology to be able to accurately, quantitatively, and systematically evaluate QPE algorithms. The extensive use of QPE data in numerical weather prediction models, for verification of quantitative precipitation forecasts, in precipitation monitoring, and as input to hydrologic models is rather diverse and needs to be evaluated.

Moreover, a consistent evaluation methodology will be vital to developers of new QPE algorithms as more remotely sensed rainfall data become available through radar and space-born technologies. The developed evaluation methodology assesses the skill of each rainfall algorithm after they have been submitted to the same degrees of model parametric uncertainty. Probabilistic measures are used to describe the accuracy of the ensemble of model predictions that were created using differing inputs. The stochastic approach proposed herein ensures that the probabilistic measures are computed independently of the inputs that comprised the calibration data set and is thus completely objective.

"Predictability" in environmental modeling encompasses a broad area of research (NRC 2001). The NRC (2001) report uses classical definitions for predictability sources; that is information present in the initial and boundary conditions. A third source of predictability is introduced as the nonlinear scale interactions in hydrologic systems. The first task in this study involves a refinement of the problem statement. In general, a modeling system possesses some degree of predictability if forecasts of a given variable have been produced and compared to observed system behavior, and the forecasts show skill over a benchmark measure such as persistence or climatology. This definition assumes that the metric used to evaluate the forecasts takes into account the *uncertainty* that is supplied with each forecast. So, what is predictability? In the context of this study, the *characteristics* of prediction uncertainty are first identified and then

quantified. Specifically, the first characteristic of prediction uncertainty in this study is learning how a perturbation in the model input propagates through the model to yield a forecast. Does this perturbation get dampened in the modeling process or does it grow?

It would be informative to meteorologists and hydrologists alike if it were discovered that significant errors in QPE resulted in negligible hydrologic prediction uncertainties. While this circumstance is unlikely, the methodology proposed herein has the capability to quantify the impact of perturbing rainfall amounts and examining their impacts on simulations of streamflow. The magnitude and behavior of the impact of these rainfall errors helps to define the needed precision and accuracy in rainfall algorithms and observations of river discharge. Moreover, this information can be used to define how much detail is needed in the physical modeling process itself. For example, many traditional hydrologic models such as the Sacramento model (Burnash 1995) rely on the principle of superposition in unit hydrograph theory. If perturbations in rainfall fields result in nonlinear basin responses, does such a model structure that relies on linearity seem valid? The method proposed herein will answer this and other related questions.

Another revealing characteristic of prediction uncertainty is the identification of the components of uncertainty and quantifying their relative contributions. Figure 3.1 shows a diagram of the simplistic model. There can be

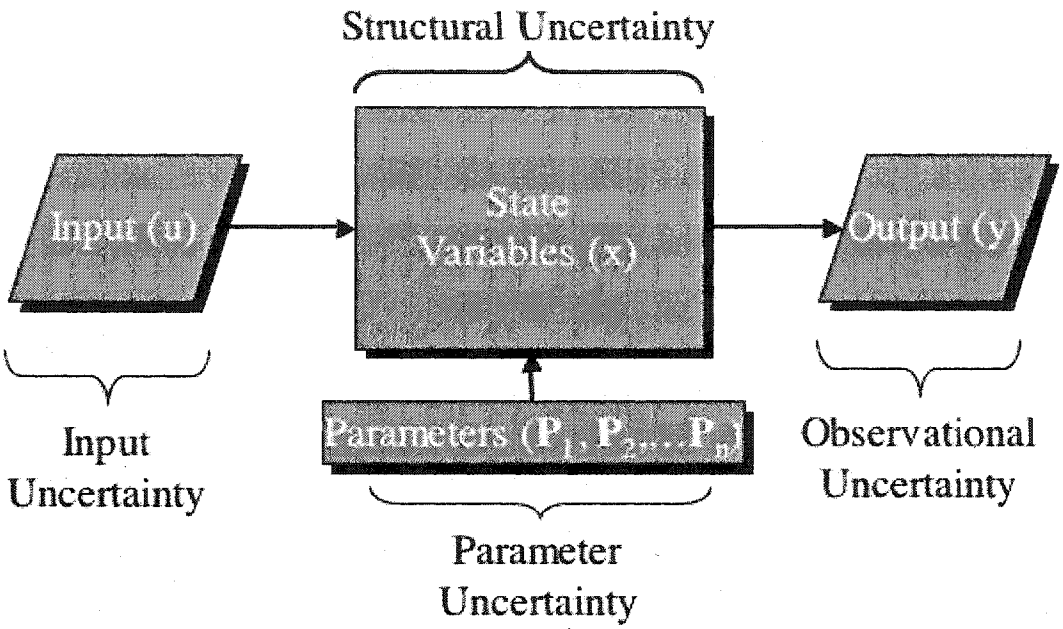


FIG. 3.1. A schematic of the components of a simple model. Uncertainty may be present in the areas noted.

uncertainty in the model inputs (e.g., rainfall estimates), the model structure (e.g., governing equations), model parameters (e.g., Manning roughness parameter derived from land use/cover maps), and observations of the system behavior (e.g., streamflow). The *total prediction uncertainty* must be estimated by incorporation of uncertainty in as many of these areas as possible, as it is the summed effect of the aforementioned uncertainty. Several methods have been devised to account for uncertainty in the model parameters (see Chapter II, section 2.3 for a complete discussion). Thus far, no known studies have computed the summed effect of uncertainty in predictions by including uncertainty in the model inputs, i.e., the rainfall. The uncertainty associated with *in-situ* observations of river discharge are not explicitly accounted for in this study; however, streamflow measurements have errors on the order of 3-6% (Sauer and Meyer 1992) and are second order as compared to the other contributions of uncertainty. Once the uncertainty due to model parameterization and input errors are quantified, it is shown how other sources of uncertainty that are more difficult to identify, such as that within the model structure itself, are elucidated.

Better estimation of the combined prediction uncertainty provides for a more realistic evaluation of the skill of the hydrologic prediction system. Moreover, the ensemble approach enables hydrologic predictions to become more useful to users of these forecasts if the total prediction uncertainty is estimated more accurately. The creation of ensembles requires running the models, or

collection of models, many times in order to encompass the phase space of possibilities. Computational constraints, although becoming less and less prevalent, engage the modeler to devise ensembles that produce the most spread while minimizing the number of simulations needed. After the uncertainties from inputs and parameters have been included in a combined ensemble prediction system, how many simulations are really required to produce the expected realm of possibilities? What is the *total* prediction uncertainty? This study addresses and answers each one of these questions.

3.2. Experimental Plan

3.2.1. Case Data

Prior to performing the hydrologic simulations, model inputs, parameter maps, and observations of output must be assembled. The *Vflo* ™ model requires the following parameters that are distributed spatially, but constant in time: saturated hydraulic conductivity (K), initial degree of soil saturation (θ), soil suction at wetting front (ψ), terrain slope (S_o), flow direction, flow accumulation, and the Manning roughness coefficient (n). In addition, the geometry (side slope and bottom width) and hydraulic roughness of the defined channels must be specified. Prior studies on the Blue River Basin have resulted in the auspicious arrangement of all the aforementioned parameter maps and channel cross-

sections. All model parameters are analyzed on a common grid. The grid cell resolution is 270x270 m.

Model inputs are supplied from the National Weather Service's Arkansas Basin River Forecast Center P1 algorithm (see Chapter II, section 2.1.4) and from the National Severe Storms Laboratory's Quantitative Precipitation Estimation and Segregation Using Multiple Sensors (QPE SUMS) algorithm (see Chapter II, section 2.1.4). The nominal resolution of the P1 products is 4000x4000 m produced hourly. The latter algorithm produces a suite of rainfall estimates, some of which rely on calibration from rain gauge measurements from the OK Mesonet (see Chapter II, Section 2.1.1). The original resolution of QPE SUMS products ranges from 5-min to 1-hour on a grid having 1000x1000 m grid cell resolution. Hourly data are used in this study as opposed to the 5-min data in order to mitigate uncertainties due to temporal differences alone. All rainfall products are resampled on the common grid employed by the *Vflo* ™ model. Hydrologic forecasts are verified using US Geological Survey (USGS) hourly discharge observations on the Blue River near Blue, OK (site ID 07332500).

The hydrologic evaluation focuses on three cases that resulted in significant flow on the Blue River. Details about each case including data availability are summarized in Table 3.1. Recently, there have been significant improvements and increased diversity of the sensors used in QPE algorithms. Rainfall data from the latest developments in QPE have been archived for a

handful of events in OK from July through December of 2002. The QPE SUMS algorithm produces the following outputs over the entire state of OK: multisensor rainfall (MS), multisensor rainfall with mean field bias adjustment using Mesonet gauges (MS-G), multisensor rainfall with local bias adjustment (MS-LG), multisensor rainfall with mean field bias removed and local bias adjustment applied (MS-LGG), radar-only rainfall (RAD), radar-only rainfall with mean field bias adjustment using Mesonet gauges (MS-G), radar-only rainfall with local bias adjustment (RAD-LG), radar-only rainfall with mean field bias removed and local bias adjustment applied (RAD-LGG) and gauge-only rainfall (GAG). Detailed explanations of each algorithm are provided in Chapter II, section 2.1.4. The developed methodology is intended to reveal the value of adding sensors to QPE algorithms such as satellite and rain gauge data. The hydrologic evaluation is also designed to identify the types of gauge adjustment strategies that lead to the most accurate hydrologic predictions. A hydrologic methodology that uses streamflow as an independent measurement is vital for evaluating QPEs that use rain gauges. Results derived from this study may be specific to the study basin, hydrologic model, and/or precipitation inputs. However, it is the methodology of evaluating the model inputs using an ensemble approach that is unique and directly applicable to other environmental models.

Several aspects of prediction uncertainty are investigated in the second phase of the study. Data from a significant cool season hydrologic event that

occurred in November of 1994 (Table 3.1) are used to determine the propagation characteristics of rainfall errors to streamflow predictions. Precipitation inputs for this particular investigation are provided by the P1 algorithm. Next, the ensemble approach is used to identify additional uncertainties in the *Vflo* ™ model formulation using an event that occurred during August of 2002. Precipitation inputs for this warm season event are supplied by the QPE SUMS algorithm. The three cases used in the hydrologic evaluation are also used to produce ensembles based on model inputs, model parameters, and a combined input-parameter ensemble. The relative skill and spread of each ensemble is quantified and compared. This information, combined with the optimized number of simulations needed to produce an accurate ensemble, provides guidance leading to the creation of ensembles useful for operational application of this hydrologic prediction system.

3.2.2. Hydrologic Evaluation using a Model Parameter Ensemble Approach

The uncertainty of hydrologic predictions cannot be assessed with a high degree of accuracy without knowing the error characteristics of the model inputs, i.e., the rainfall. New remote sensing technologies from ground-based radar, space-borne radar, and satellite are now providing rainfall estimates at scales on the order of tens of meters every few minutes. The difficulty in measuring the improvement and impact of these technological advances is the lack of

information regarding the ground truth. Traditionally, this has been accomplished using rain gauges. These *in-situ* measurements have their own errors, perhaps comparable to those of the remotely sensed data, and are sensing rainfall at a much smaller scale. For these reasons, a new approach must be considered to quantify the uncertainties in the rainfall algorithms and be performed at the hydrologic scale of application.

Prior to performing any hydrologic analysis, the hourly QPEs from each algorithm (refer to Table 3.1) are aggregated in time to yield storm total precipitation. These storm total plots are shown for each case in order to reveal subjective differences in the estimates. While this brief analysis fails to quantify the true hydrologic uncertainties with the QPEs, it does reveal characteristics associated with each algorithm. For example, the output from the GAG product is expected to take on a much different appearance from a RAD product. Moreover, artifacts due to remote sensing of variables indirectly related to rainfall (e.g., radar reflectivity) are revealed. The storm total plots show differences in the spatial distributions and magnitudes of rainfall estimated by each algorithm.

The designed hydrologic evaluation methodology must be objective and unbiased towards a given QPE input. Global optimization methods in model calibration seek a unique parameter set that best simulates the observed behavior (see Chapter II, section 2.3 for a complete discussion). If another input is used, then the model would likely need to be recalibrated as the previous parameters

were likely tuned to the expected rainfall inputs. The evaluation methodology proposed herein computes probabilities by examination of the allowable parameter space for each QPE independently and thus remains unbiased towards a given QPE. Model parameter ensembles are created for each input independently. The spread and accuracy of the *compilation* of individual simulations are determined based on comparisons with observed streamflow. It is hoped that this evaluation methodology will be adopted for systematic evaluation of new QPE algorithms.

The first step in setting up a model parameter ensemble involves assigning the ranges and distributions of parameter values. If no information is known *a priori* about the parameter distributions, then a uniform distribution can and will be assumed. The *Vflo* ™ model (see Chapter II, section 2.2) utilizes maps of saturated hydraulic conductivity (K), initial fractional water content of the soils (θ), soil suction at wetting front (ψ), bed slope (S_o), flow direction, flow accumulation, and Manning roughness coefficient (n) that are distributed spatially. The DEM-derived parameters, i.e., S_o, flow direction and accumulation, are well-defined, and variations of these parameters are related to scale issues that are beyond the scope of this study. The ψ parameter is inversely proportional to K as revealed in Table 4.3.1 of Chow et al. (1988). Varying the ψ parameter in addition to the K parameter increases the dimensionality of the parameter space, thus increasing the number of simulations by an order of magnitude. For these

reasons, the K, θ, and n parameters are assigned ranges while the other parameters remain fixed at their predefined values.

The K, θ, and n parameters are distributed in space and are derived from ancillary data. The inference of K and n from soil types and land use/cover data is associated with some uncertainty because they are not measured directly. For this reason, they must be perturbed within their physical bounds. Perturbing spatially distributed parameter maps can be a daunting task if one were to consider perturbing the values from cell-to-cell. The parameter space would increase beyond bounds that are not practical to sample and perform model simulations. The Ordered Physics-based Parameter Adjustment (OPPA) method described by Vieux and Moreda (2003) employs scalars to adjust parameter maps so that the magnitudes change while the spatial variation is preserved. This method maintains spatial variability commensurate with distributed parameter modeling while minimizing the dimensionality of the parameter space. The scalars used to multiply the K and n parameter maps are defined as follows:

$$N_i = \frac{1}{8}(2+3i)\Big|_{i=0,1,2,3,4}, \tag{3.1}$$

where N_i is the adjustment factor. In essence, (3.1) employs scalars that range from quartering to nearly doubling the parameters from their given values.

Perturbing the Manning roughness coefficients results in smooth (rough) surfaces that have the affect of reducing (increasing) the amount of time it takes water moving overland to reach the basin outlet. Increasing (decreasing) the saturated hydraulic conductivity has the affect of increasing (decreasing) potential infiltration rates (2.12). In this case, more (less) water is lost to the soils and a lower (higher) volume of water reaching the basin outlet results. The initial soil saturation parameter is varied from 20% (dry) to 100% (saturated) in increments of 20%. The *Vflo* ™ model is run on an event-based mode, thus there is no known information about spatially distributed antecedent soil moisture conditions. The parameter is essentially varied from dry to moist conditions.

The next procedure in setting up a model parameter ensemble involves running multiple hydrologic simulations using combinations of possible parameter settings for each QPE input. If there were information available regarding the initial parameter distributions, then an algorithmic procedure would be devised to sample the parameter space strategically to avoid excessive computational expense. A uniform distribution is assumed here, and the parameter space is sampled thoroughly such that no particular parameter setting is favored and given more weight in computing the final probabilities.

The first simulation uses rainfall inputs from the RAD product for the 23 October 2002 event (Table 3.1). Scalars used to multiply the n and K maps are determined by (3.1), while the first parameter setting of θ is set to 20%. Results

from this simulation are stored in a file. This procedure is repeated over and over until all possible parameter combinations have been utilized in the model. Since there are five different scalars used to adjust each of the three parameters, a total of 125 simulations are performed for the RAD rainfall input for the first event. Next, the RAD-G product is input to the *Vflo* ™ model in lieu of the RAD product. The exact same parameters are iterated through the model, with the only difference being the model inputs. Again, model simulations using the RAD-G inputs are stored in a separate file. This procedure is repeated for all available rainfall inputs for the 23 October 2002, 28 October 2002, and 03 December 2002 events (see Table 3.1).

Results from each simulation are compared to observed streamflow in the verification step. There are several ways in which predictions of streamflow can be evaluated based on comparison to observations. In the hydrologic literature (see e.g., Beven and Binley 1992; Romanowicz et al. 1994), a particular model's goodness-of-fit is typically determined using objective functions that rely on the sum of squared errors. An individual error results from the difference between discharge predictions and observations at each observational time step. The error variance is then computed throughout an event or during an entire season. The objective functions described above have useful statistical properties, but can result in a loss of information about a given hydrograph. Consider for example a predicted hydrograph that mirrors the observed perfectly, but is offset in time by a

few hours. The errors computed at each observational time step would be significant in this case, as would be the error variance throughout the event. The objective functions described above would have little capability in differentiating this kind of a predicted hydrograph versus another that has a significantly different shape. Similar magnitudes in error variances can be achieved from predicted hydrographs that have much different shapes and behavior. This study utilizes three objective functions to describe the degree to which each model matches streamflow observations. These objectives provide more information about hydrographs shapes than statistics based on error variances.

The following variables are computed from the observed discharge and each simulation of streamflow: integrated discharge throughout the storm event normalized by the basin area (Volume hereafter), maximum discharge throughout the event (Peak hereafter), and the time at which the maximum discharge occurred (Time hereafter). Figure 3.2 shows how each of these variables are derived from a given hydrograph. 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 (see Fig. 1.1). The extent to which each product may possess biases, especially at medium

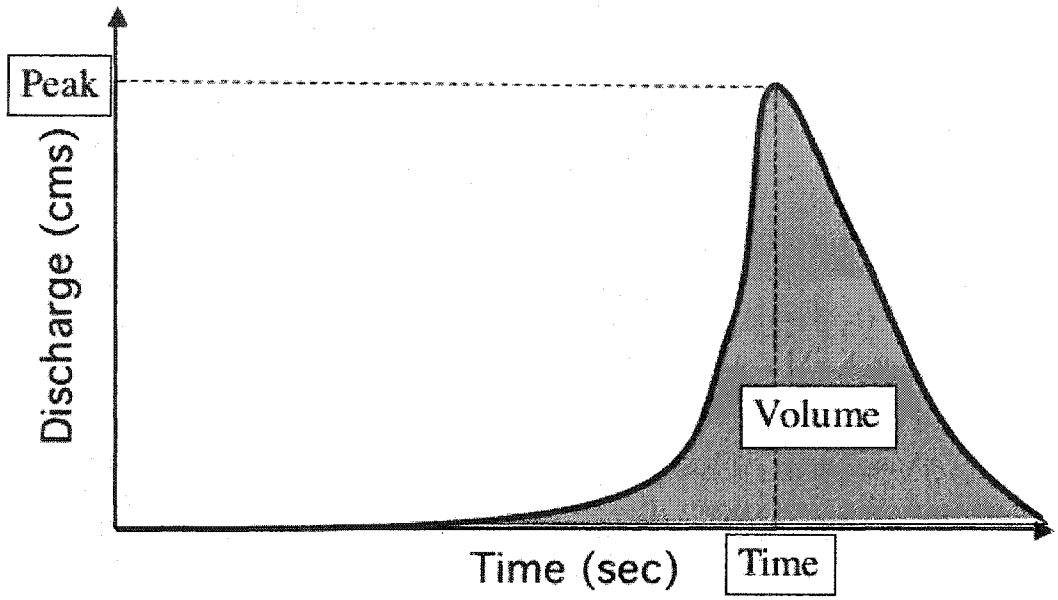


FIG. 3.2. A schematic of a typical hydrograph and the derived variables that are used in this study. The Time (sec) variable is the time at which the maximum discharge occurs, Peak (cms) is the magnitude of the maximum discharge, and Volume is the time-integrated flow under the hydrograph that's been normalized by the basin area (mm).

to far ranges from radar, 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.

The simplest procedure to verify an ensemble of forecasts is to derive and evaluate the first moment of the forecast probability distribution, the mean. In numerical weather prediction, it has been shown that the mean of the forecast ensemble will have a smaller error than the mean error of individual forecasts (Leith 1974; Murphy 1988). Thus, the mean of the forecasts, when compared to observations, provides an overall indicator of the ensembles' behavior. This evaluation methodology lacks robustness because it is assumed *a priori* that the forecast probability distributions can be approximated by a Gaussian distribution. These parametric statistics are attractive due to their simplicity, but additional information about the entire probability distribution functions (pdfs) is revealed using statistics developed in following sections.

The mean Time, Peak, and Volume of each ensemble is computed. The following statistics are utilized in order to compare the ensemble averages to observations: *Bias*, Mean Absolute Error (*MAE*), and root mean squared error (*RMSE*). Their definitions are provided below:

$$Bias = \frac{E[F_i]}{O}, \tag{3.2}$$

$$MAE = E\left[\left|F_i - O\right|\right], \tag{3.3}$$

$$RMSE = \sqrt{E\left[(F_i - O)^2\right]}, \tag{3.4}$$

where F_i represents a forecast from the ith simulation for Time, Peak, and Volume, O is the observation, and $E[\ ]$ is the expected, or mean value. *Bias* shows how ensemble means compare to observations in an overall sense. *MAE* and *RMSE* reveal the degree of scatter or variability between individual forecasts and observations. *Bias*es closest to 1, and *MAE*s and *RMSE*s closest to 0 by definition have the best agreement with observations of streamflow. Observed rainfall is compared to observed streamflow using a runoff coefficient (RO Coeff hereafter). This quantity is the discharge volume divided by the basin-averaged precipitation summed throughout each event.

The statistics described in (3.2-3.4) are simplistic and provide a first order evaluation of the hydrologic variables as compared to observations. However, they only consider the ensemble mean and not the full spread and distribution of predictions. A forecast ensemble provides the *capability* to estimate the forecast pdf of the hydrologic variables. A more robust statistic is developed below to compare the predicted pdfs to a single observation. In this hydrologic evaluation the forecast "probability" distribution is a result of parametric uncertainty alone. It is possible, but unlikely, that errors in the model structure and/or observations

of streamflow are structured to favor a given model input. The word "probability" is used cautiously here because the distributions of forecast variables may not necessarily be representative of the "true" probability distribution function. The probability is more aptly referred to as a conditional probability. This concept is explored further in this Chapter.

Probability distribution functions are computed for Time, Peak, and Volume using each 125-member ensemble representing the different QPE inputs. In many applications, a histogram would be a sufficient qualitative manner in which to display the distribution of a given dataset. However, it must be noted that a judicious choice of the number of classes used and the class interval can affect the appearance and conclusions drawn from the computed histograms. Moreover, histograms are not always smooth or continuous which reduces their utility if derivatives are needed. This study computes pdfs for each ensemble using Gaussian kernel density estimation (Silverman 1986). In a simplified manner, this density estimation technique can be thought of as a smoothed histogram. The resulting function is smooth and continuous which leads to better estimation of probability exceedence (i.e., 90% simulation limits) and calculation of derivatives.

Following Silverman (1986), a kernel density estimate is computed as follows:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{3.5}$$

where $\hat{f}(x)$ is an estimate of the data density, h is the smoothing parameter or bandwidth, K is the kernel, X_i is the ith prediction, and n is the number of ensemble members (with each supplying a prediction). The kernel is actually a function itself which satisfies the following condition:

$$\int_{-\infty}^{\infty} K(x)dx = 1. \tag{3.6}$$

In this case, the kernel is nonnegative and from (3.6) satisfies the requirements for a pdf. In (3.5), note how the pdf is computed from the sum of individual kernels at the prediction locations. In this sense, the estimated pdf is a compilation of several "bumps", where the shape of the bump is defined by the kernel estimator being used. In this study, a Gaussian kernel is applied. Thus the kernel density estimate becomes

$$\hat{f}(x) = \frac{1}{\sigma\sqrt{2\pi}}\sum_{i=1}^{n} e^{-\left\{\frac{1}{2}\left[\frac{(X_i - x)}{\sigma}\right]^2\right\}}, \tag{3.7}$$

where σ is the bandwidth. The choice of the bandwidth defines how smooth the estimated pdf will be. Smaller (larger) bandwidths lead to more (less) bumps in the estimated pdf that may or may not be real. In essence, small (large) bandwidths place lower (higher) weight on observations that are further away. If only one pdf were being estimated, then the bandwidth could be chosen subjectively to minimize the insignificant bumps while retaining the "true" shape of the pdf. In this study, many pdfs are estimated requiring an objective choice of the smoothing parameter, σ. Silverman (1986) recommends using a bandwidth that is adaptive to the range or spread of predictions. The following equation is used to adaptively adjust the bandwidth to accommodate each dataset:

$$\sigma = 0.79Rn^{-\frac{1}{5}}, \tag{3.8}$$

where R is the interquartile range of the predictions. Plots of the pdfs are produced and subjectively compared. Additional statistics are derived from the density estimates.

It is useful to the modeler to understand performance measures of a hydrologic model using a given rainfall input. Specifically, measures of central tendency, spread, and skill are computed from each pdf. A nonparametric measure of central tendency is the 50% simulation limit, or median. This is the

value corresponding to 0.5 on the cumulative distribution function (cdf). Similarly, the spread can be quantified by determining the distance between the 5% and 95% simulation bounds. Similar to the median, these simulation limits are determined from the values corresponding to 0.05 and 0.95 on the cdf. This provides for measures of central tendency and uncertainty limits for each ensemble. Plots from these measures are produced for each ensemble, thus enabling subjective comparison of the simulation limits.

It would be informative to the modeler to know the ensemble skill based on the pdfs. This approach is nonparametric and considers each member with equal weight. Statistics used to evaluate probability forecasts for multicategory events have been developed. A condition for a probability distribution is that the probability for a given event must be greater than or equal to zero and less than or equal to one. Secondly, the sum of probabilities for the event must be one. The pdfs computed from the ensembles may not be an accurate estimate of the true underlying pdf if there are uncertainties in the model structure, model inputs, or observations of model predictands. Thus, the pdfs computed here are treated as conditional probability distributions. Nonetheless, ensemble forecasts are verified using the entire data distribution with the understanding that they are most likely not accurate estimates of the true pdf.

The ensemble skill is assessed using the ranked probability score (*RPS*; Wilks 1995). This statistic is particularly useful for probability forecasts of an

event that has a natural ordering. Consider an example where a given ensemble forecasts an observed Volume of 6 mm with only a 10% probability, while another ensemble forecasts the same observation also with 10% probability. If we were only considering one category, then these forecasts would have the same skill. However, the *RPS* accommodates multi-category events so that the probability assigned to the other categories is assessed. Using the same example, the first example might forecast a Volume in an adjacent category of 7 mm (i.e., only 1 mm different from the observed) with a 90% probability, while the second ensemble forecasts a Volume of 12 mm with a 90% probability. In this case, the second ensemble is penalized more heavily because more probability was assigned to a category further removed from the observation. Thus, the *RPS* is said to be sensitive to distance, an attractive property to the skill score. The formal definition if provided below:

$$RPS = \sum_{m=1}^{J}\left[\left(\sum_{i=1}^{m} y_i\right) - \left(\sum_{i=1}^{m} o_i\right)\right]^2, \qquad (3.9)$$

where y_i is the cumulative probability assigned to the ith category, o_i is the cumulative probability of the observation in the ith category, and J is the number of categories. The *RPS* values require the selection of categories. Ten categories were chosen based on observed Time, Peak, and Volume observations. The first

Time category was chosen to be the observed time of maximum discharge minus one day, with a class interval of five hours. The first Peak and Volume categories were chosen to be 25% of the observed variables with class intervals equal to the first category. The choice of categories is a subjective one, but is based entirely on observed data, not on the predictions. This enables the *RPS* scores to be compared to one another in an objective way.

Probability distributions and subsequent calculation of *RPS* values from two independent ensembles may certainly appear to be different. However, in order to justify the statistical significance of the differences, a resampling technique is employed. The resampling technique pools together all of the ensemble members from two different ensembles. Two new ensembles of the same size are created by randomly selecting members from the pooled population. *RPS* values are computed for the new ensembles and are stored. This procedure is repeated 1000 times. Each *RPS difference* is computed and used to produce a cumulative distribution. The cumulative distribution is then used to determine the probability of obtaining the original *RPS* difference. These confidence levels are computed for each ensemble that is being compared. For the 23 October 2002 event, there are as many as nine different model parameter ensembles representing nine different model inputs (see Table 3.1). The confidence intervals are thus presented in a 9x9 matrix where repeated results are omitted.

A hydrologic evaluation of QPE algorithms is accomplished through the development of ensembles whose skill is measured by the *RPS* score. These ensembles are produced for each QPE input independently. Each parameter is varied within its physical bounds using scalars to produce model parameter ensembles. First order parametric statistics such as the *Bias*, *MAE*, and *RMSE* are used here to compare the ensemble mean to the observations of streamflow. A major benefit of ensemble forecasting is the possibility of estimating the true forecast probability distribution. In the hydrologic evaluation method proposed herein, pdfs are computed for each QPE input. As discussed previously, the word "probability" is used here with the understanding that it is a probability conditioned on perfect model physics, perfect observations of streamflow, and accurate estimates of rainfall, all of which are rarely obtainable. The pdfs are compared in order to determine the rainfall input that results in the most likely forecasts of streamflow, still conditioned on the perfect model and perfect observations assumption. *RPS* values are compared, and the statistical significance of the differences is evaluated. The parameter ensembles also enable the computation of simulation limits, or 5%, 50%, and 95% quantiles. These derived statistics indicate the spread or uncertainty with the predicted ensemble and a measure of central tendency. The combination of these analyses reveals the model input that has the best chance of producing hydrologic forecasts, i.e. best simulates the observed system behavior.

3.2.3. Evaluation of the Characteristics of Prediction Uncertainty

3.2.3.1. Propagation Characteristics of Input Uncertainty

The third objective in this study is to determine the sensitivity of simulations of streamflow to errors in rainfall measurements. The assumption of linearity in the transformation of rainfall to runoff is tested upon examination of the impact of perturbing rainfall amounts. If increasing the rainfall inputs to the model by a factor of two results in a similar error in simulated discharge, then the assumption of linearity is vindicated. On the other hand, perhaps a large perturbation in rainfall results in an error in predicted discharge that has a smaller relative magnitude. This information indicates that the model predictions are not sensitive to the QPE inputs. Conversely, perhaps the introduced errors in rainfall inputs blow up into much greater errors in discharge. This nonlinearity in the relationship between rainfall and runoff would have large impacts on current models that utilize unit hydrograph theory. Moreover, the need to improve the precipitation inputs to the hydrologic model would be well justified.

This third science question is addressed using a similar methodology developed in the hydrologic evaluation described in section 3.2.2 of this Chapter. Specifically, the uncertainties in the model parameters are used to create forecast ensembles. It can be argued that the sensitivity of forecasts to input perturbations could be determined using a single model structure and parameter set. It is unlikely, but nonetheless possible, that the chosen parameter set results in an

aberrant behavior as compared to the rest of the ensemble members. For instance, if an unrepresentative, low value for initial soil saturation is used, there may not be much runoff generation even if the rainfall amounts are doubled. One might jump to the conclusion that simulated hydrograph responses are insensitive to rainfall perturbations. A model parameter ensemble is utilized here to avoid making conclusions based on deterministic forecasts, but rather the entire range of forecasts. The model inputs used to generate the ensembles are not independent of one another, but are simply perturbations of the given, deterministic rainfall input. Analogous to the perturbations imposed on the parameter maps, rainfall estimates are perturbed from one-half to double their initial value. Five scalars are used to perturb the rainfall estimates: 0.5, 0.75, 1.0, 1.33, and 2.0. The 12 November 1994 event is used for this sensitivity test (Table 3.1).

The scalars used to adjust the n, K, and θ parameters are the same ones used in the hydrologic evaluation. An ensemble is created for each of the rainfall inputs combined with 125 different configurations of model parameters, totaling five forecast ensembles. Each of the ensembles uses the same parameter settings, but each uses a different model input. After all the simulations are performed, the Time, Peak, and Volume hydrologic variables are computed from each member. The *bias*, *MAE*, and *RMSE* statistics (3.2-3.4) quantify the agreement between the forecast hydrologic variables and observations. In addition, pdfs are computed from the 125-member ensembles using the Gaussian kernel density estimation

technique. These pdfs are conditioned on an accurate model structure and observations of streamflow. A change in the model structure can certainly result in a completely different hydrologic response to the model inputs and their perturbations. This sensitivity test is thus specific to a single model structure.

The estimation of pdfs enables the computation of .05, .50, and .95 quantiles. These nonparametric statistics provide information regarding central tendency and the spread of the ensemble. The skill of each ensemble is assessed using the *RPS*. Since *RPS* values from different ensembles are being compared, the statistical significance must be evaluated using a resampling test. All of the discussed statistics are tabulated in a table and some are charted in order to reveal characteristics of input uncertainty on hydrologic forecasts. The forecast ensemble reveals the sensitivity of model forecasts to rainfall inputs by taking into account the deviations that result from all physically realistic parameter sets.

3.2.3.2. Identification of Uncertainty in Model Physics

In environmental modeling, uncertainty is known to reside in observations, model inputs, boundary conditions, parameters, and in the model structure. A forecast ensemble can be carefully crafted in order to encompass the uncertainty in all the above areas, thus providing an accurate estimate of the true pdf. Several sampling strategies of the parameter space have been developed as long as the physical ranges of each parameter can be specified (see e.g., Chapter

II, section 2.3). This is much more easily accomplished with a model whose structure and parameters have a physical basis, such as with Class I-II models (see Chapter II, section 2.2 for a complete discussion). The *Vflo*™ model uses physically based parameters, which means realistic ranges and thus perturbation scalars can be assigned a priori.

It is much more challenging to perturb the model inputs, i.e. rainfall estimates, without taking into consideration several aspects of the instrument used to measure the rainfall. For instance, rainfall estimates as measured by radar are known to have system-wide biases, as well as biases that are spatially and temporally variable. The number of simulations needed to encompass each and every one of these possibilities at each grid cell for each time step would result in an unrealistically large ensemble. This obstacle is addressed using all the rainfall possibilities from the QPE SUMS precipitation algorithm. This algorithm produces a suite of products every hour based on individual sensors and combinations. In essence, the suite of products as a whole can be treated as an ensemble of model inputs. The large number of members and diversity of instruments used to produce the QPEs are assumed, but have not been proven, to be an accurate estimate of the true, unknown rainfall pdf. Because the pdf is not known perfectly, each QPE SUMS member will be assumed to be an equally probable estimator of the true rainfall and will be given equal weight.

Forecast ensembles can now be created from the range of possibilities suggested by the combination of model inputs, model parameters, and the assumed uniformity of their distributions. At this point, an estimate of uncertainty can be obtained for a given hydrologic prediction as long as the model structure and observations of streamflow are perfect. Of course, neither of these conditions holds with certitude for all conditions. Consider, for example, potential errors in streamflow observation during low-flow conditions in a large riverbed. The water pathways can meander within the riverbed giving it a braided characteristic. A point measurement of streamflow may be an underestimation of the total flux of water leaving a given cross-section. Here, the observation of streamflow is not perfect, as is typically the case with measurements of the environment. Nonetheless, measuring streamflow at a given cross-section in a river channel is much simpler and easier to accomplish as compared to capturing the spatial and temporal variability of rainfall over an entire basin. Similarly, estimation of model parameters over the basin based on indirect measurements is also much more challenging. The uncertainties in streamflow measurements have been estimated to be 3-6% (Sauer and Meyer 1992). These uncertainties are assumed to be small as compared to uncertainties associated with model inputs and parameters. Uncertainties with streamflow measurements will be neglected in order to isolate uncertainties in other components of the modeling process (e.g., in the model structure).

It is now assumed that the uncertainties in rainfall estimates and model parameters are accounted for by accurate estimation of their respective pdfs. In addition, uncertainties caused by observations of streamflow can and are assumed to be negligible relative to the aforementioned sources of uncertainty. The remaining source of uncertainty in hydrologic forecasts rests in the model structure itself. Model structural uncertainties can result from one or more of the following: numerical errors, incomplete representations of all the physical processes important to the variable being predicted, nonrepresentative model formulations due to an incomplete understanding of the physics involved (e.g., runoff generation from Hortonian versus Dunne-type), and application of these physics to the precise scale of the basin. These unknowns are impossible to quantify for any particular environmental modeling system, otherwise the model would be perfect. The identification of model uncertainty is thus accomplished by quantification of all other error sources, with the model uncertainty being the remaining component of the total error that isn't explicitly accounted for. The null hypothesis in this case states that the model is perfect. If the null hypothesis is correct, then there are no residual errors after the model input and parameter uncertainties have been accurately accounted for. The null hypothesis may be refuted in the event that there is some residual error for the case event being studied. This hypothesis testing assumes that the sensitive parameters have been

identified, their ranges have been properly assigned, and the QPE inputs estimate the underlying rainfall pdf.

The methodology for model uncertainty identification begins by creating an ensemble that is comprised of uncertainties in the model inputs and parameters. The 25 August 2002 event is used for this experiment (see Table 3.1). The ensemble is composed of all available model inputs from QPE SUMS for this case combined with parameter perturbations used in sections 3.2.2 and 3.2.3.1. Similar to prior analyses, the *bias*, *MAE*, and *RMSE* statistics are used to verify the forecasts of streamflow. In addition, pdfs are derived from the ensemble, thus enabling the computation of simulation quantiles and an *RPS*. Confidence levels are not required in this case because *RPS* values are not being compared to one another.

The null hypothesis stating that the model structure is perfect can be refuted if the following conditions are satisfied: parameter ranges have been adequately assigned, perturbations of QPE inputs represent the true rainfall pdf, and statistics indicate that the forecast ensemble does not give the observed system behavior. Specifically, if the observations of streamflow fall outside the simulation bounds, then the residual error may be attributed to an erroneous model structure. It is also plausible that parameter ranges were not assigned to be wide enough, or all the QPE inputs are biased so that their pdf does not adequately encompass the true rainfall. A possible explanation for the model

structural uncertainty is provided and supported by evidence from independent observations of rainfall, runoff, and soil moisture conditions. Conclusions from this experiment may be specific to this particular study event, basin, and specific ranges of parameters explored. Nonetheless, the focus here is on the *technique* used to identify model structural uncertainty.

3.2.3.3. Estimation of Combined Prediction Uncertainty

A major goal in hydrologic forecasting is to estimate the total uncertainty associated with a given prediction. It is shown in section 3.2.3.2 how an ensemble can be constructed to encompass the uncertainty associated with model inputs and model parameters, two major sources of uncertainty in hydrologic prediction systems. This quantification of uncertainty leads to the *probability* that a certain hydrologic condition will be accurately predicted given hydrometeorological inputs. The other sources of error lie in the model structure and observations of streamflow. The latter potential error source has been quantified and is negligible in most cases in comparison with the other uncertainties. Model structural errors with the *Vflo* ™ model have been discovered, but will be shown to be isolated to specific times of the year. These model structural errors are not as prevalent for the three cases used in the hydrologic evaluation (section 3.2.2). The total prediction uncertainty can thus be estimated by incorporating the uncertainties inherent in model inputs and

parameters into a combined ensemble. Probabilities of hydrologic conditions are accurately estimated for the same three cases used in the hydrologic evaluation (refer to Table 3.1).

A combined ensemble is constructed as described previously by estimating the rainfall pdf and the pdfs for each model parameter. The rainfall pdf is estimated by treating each QPE SUMS member as if it is an equally likely estimator of the true rainfall field. This is a valid assumption because the QPE SUMS products use all available rainfall sensors and logical combinations in their estimation schemes. A uniform distribution of the rainfall products will be assumed here. Similarly, the three sensitive parameters used in the *Vflo* ™ model are perturbed within their physical bounds as in (3.1). No assumptions are made a priori about the true underlying pdfs of the parameters. Thus, a uniform distribution is assumed for the initial parameter settings and their perturbations, i.e., they are each given equal weight. Future research will reject those models (e.g., given rainfall input and parameter set) that give a prediction that is deemed nonbehavioral. This will further refine the simulation bounds for hydrologic predictions on the Blue River Basin.

The number of members comprising a combined ensemble is a function of the available rainfall inputs for a given case (see Table 3.1) multiplied by the number of parameter possibilities (125). After the combined ensemble is constructed, a pdf is estimated using the Gaussian kernel density estimation

technique and plotted for each predicted variable. The *bias*, *MAE*, and *RMSE* statistics (3.2-3.4) are computed from the ensemble mean and placed in a table. The nonparametric *RPS* is computed to determine the skill of the hydrologic prediction system and 90% simulation limits are also computed from the pdfs. The simulation limits are plotted along with the observed variables to show the degree of spread or uncertainty associated with each hydrologic prediction. This prediction uncertainty analysis is performed for each hydrologic variable (i.e., Time, Peak, and Volume) for all three cases.

3.2.3.4. *Optimum Number of Ensemble Members*

The previous section demonstrates how an ensemble can be constructed to include the dominant sources of uncertainty for the cases studied. Probabilistic forecasts that a given hydrologic variable will be met or exceeded can be estimated with this technique. These statistics have large potential value when used in practice. Consider, for example, a drainage that is near a residential neighborhood. Let's say that flood stage is reached when the maximum discharge exceeds 100 cms. Emergency management officials may establish a threshold probability (e.g., 40%) that needs to be exceeded in order to evacuate the neighborhood. The developed ensemble prediction system is capable of estimating these probabilities. When considering the application of such a prediction system in real-time, one must be cognizant of potential computation

constraints when utilizing hundreds of model simulations. A methodology has been devised to evaluate the skill of an ensemble of predictions as a function of the number of members comprising the ensemble. If the number of members can be reduced, then the number of simulations and thus computational requirements can be minimized. The approach taken is similar to that shown in Du et al. (1997). The 23 October 2002, 28 October 2002, and 03 December 2002 events (Table 3.1) are used once again for this test.

Ensembles are first constructed using the maximum number of parameter settings (125) coupled with the number of available rainfall products for the three cases used in section 3.2.2 and 3.2.3.3 (refer to Table 3.1). Thus, a combined ensemble for the 23 October 2002 event has a maximum number of 1125 members. Monte Carlo permutation techniques are used to randomly sample individual members from the total ensemble. The number of members sampled ranges from 1 to the maximum number of members comprising the ensemble. This procedure is performed 1000 times and the average *RPS* value is reported for the number of members used in each ensemble. The repetition and subsequent averaging was necessary to ensure that the smaller ensembles didn't cluster in a favorable or equally unfavorable region of the phase space. Plots are created that show the sensitivity of the *RPS* to the ensemble size. Moreover, the number of ensemble members needed to achieve 90% of the skill attained when considering all members is reported.

The results reported in the next Chapter may indeed be specific to the *Vflo* ™ model, the QPE SUMS rainfall inputs, the events being examined, and/or application of the model on the Blue River Basin. Nonetheless, the focus of this study is on the methodology of utilizing an ensemble prediction system to evaluate the hydrologic performance of different rainfall products in a quantitative, objective manner, a major goal in hydrometeorology. The second major undertaking is the evaluation of several characteristics of prediction uncertainty. Specifically, the approach taken herein will reveal the sensitivity of streamflow predictions to perturbed rainfall inputs, isolate and identify model structural errors, provide an estimate of the total prediction uncertainty, and optimize the number of ensemble members needed for accurate estimation of the forecast probabilities.

TABLE 3.1. A summary of the hydrologic events studied. The Observed Time refers to the time of peak discharge, the Observed Peak is the magnitude of the peak discharge, and the Observed Volume is the time-integrated discharge normalized by the basin area. See Chapter II, section 2.1.4 for a complete description of the rainfall products.

Case Start Time	Case End Time	Observed Time	Observed Peak (cms)	Observed Volume (mm)	Available Rainfall Products
0000 UTC 11/12/94	0000 UTC 11/24/94	2000 UTC 11/15/94	215.2	41.1	ABRFC P1
0000 UTC 08/25/02	0000 UTC 09/01/02	0900 UTC 8/28/02	6.2	1.6	GAG, RAD, RAD-G, RAD-LG, MS, MS-G, MS-LG
0000 UTC 10/23/02	0000 UTC 10/30/02	0000 UTC 10/26/02	56.7	6.1	GAG, RAD, RAD-G, RAD-LG, RAD-LGG, MS, MS-G, MS-LG, MS-LGG
0000 UTC 10/28/02	0000 UTC 11/01/02	0600 UTC 10/30/02	16.4	2.3	GAG, MS, MS-G, MS-LG, MS-LGG

0000 UTC 12/03/02	0000 UTC 12/10/02	1600 UTC 12/05/02	13.4	2.5	GAG, RAD, RAD-G, RAD-LG, RAD-LGG

CHAPTER IV. RESULTS

4.1. Hydrologic Evaluation of Quantitative Precipitation Inputs

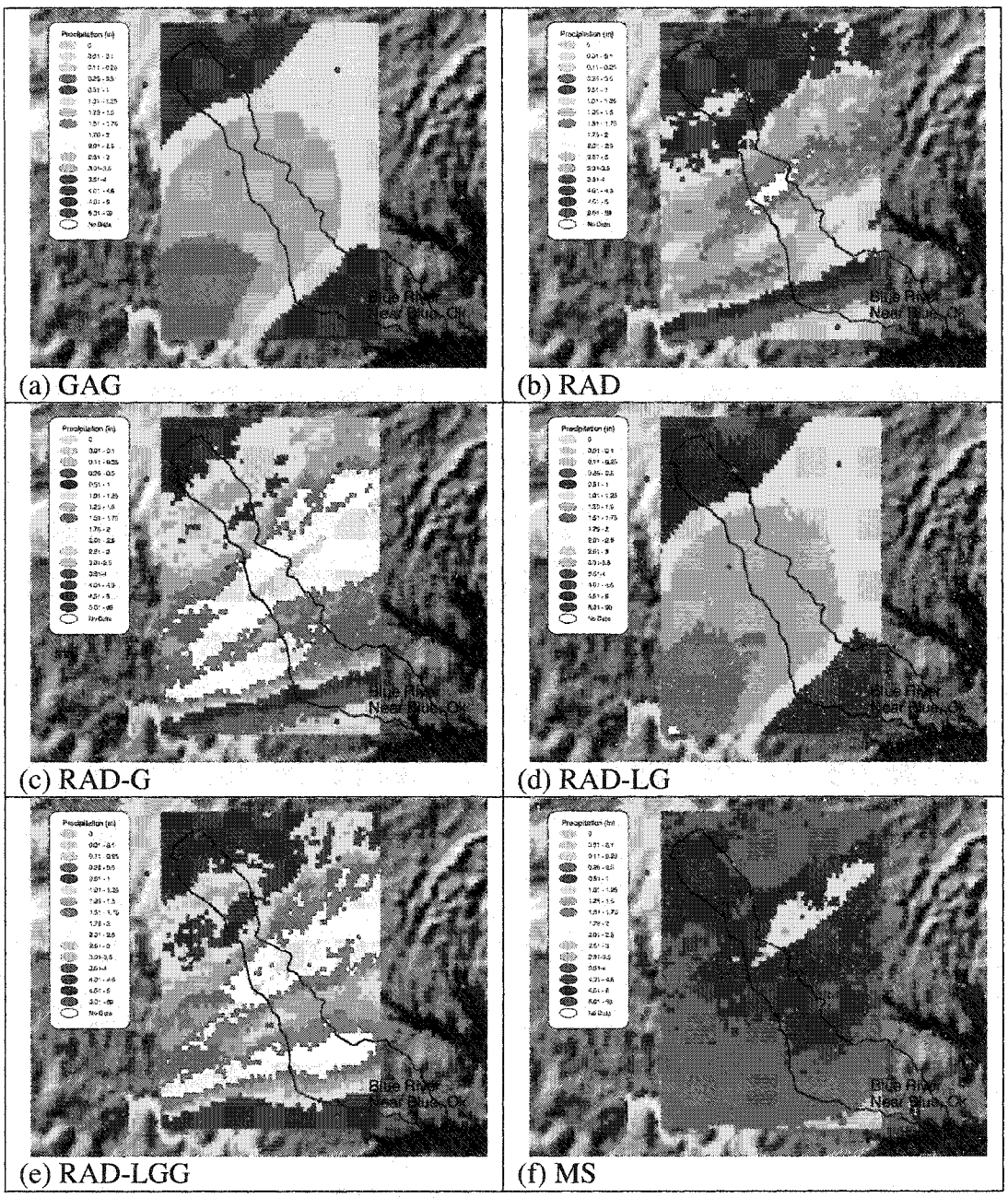
Three precipitation events during the fall of 2002 (Table 3.1) resulted in significant hydrologic responses on the Blue River Basin in OK. A unique methodology has been developed to evaluate the relative skill of hydrologic predictions using different quantitative precipitation estimates (QPEs). Traditionally, remotely sensed QPEs are compared against independent rain gauge observations. In this study, the evaluation is performed at the scale of application, an integrating basin. The results from this study reveal the estimator that is most likely to provide the best input to the *Vflo* ™ model. These results are significant because the main impetus of improving QPE algorithms is to ultimately improve hydrologic predictions. This study will highlight those rainfall algorithms that are most capable of predicting streamflow, and thus mitigating the impacts of flash flooding and river flooding. Moreover, the methodology requires no prior modeling experience on a basin (i.e., calibration) and is completely objective.

4.1.1. 23 October 2002 Case

A subjective comparison of the rainfall products is accomplished by simply aggregating the hourly accumulations over the duration of the entire storm.

Storm total plots for the 23 October 2002 case are shown in Figs. 4.1a-i. Visual differences are apparent between the panels indicating the dependence of the rainfall patterns on the sensor(s) used to produce the estimates. For example, the gauge-only product (GAG; Fig. 4.1a) is quite smooth in comparison with the radar-only (RAD; Fig. 4.1b) product. The gauge-only product is composed of point estimates of rainfall that have been analyzed on a Cartesian grid. The parameters used in the analysis scheme dictate the degree of smoothness in the final products. The radar-only product, on the other hand, collects independent measurements of rainfall at each grid point. The spatial and temporal resolution of the radar data is observably superior, however the radar doesn't measure rainfall directly like the rain gauges. The multisensor product (MS; Fig. 4.1f) bears some resemblance to the radar-only product in terms of the maximum rainfall. However, the MS product is noticeably smoother and reflects the use of satellite data in its scheme.

Three different rain gauge adjustment strategies are being evaluated in this study. Figures 4.1c-e show how these impact the RAD product. Mean field bias adjustments applied to the RAD field (RAD-G; Fig. 4.1c) maintain the spatial structure in the original product while the magnitude has been scaled. In this case, the adjustment procedure results in heavier accumulations of rainfall throughout the storm duration. The multisensor product with mean field bias adjustment (MS-G; Fig. 4.1g) is affected in a similar manner. The local bias



(a) GAG (b) RAD

(c) RAD-G (d) RAD-LG

(e) RAD-LGG (f) MS

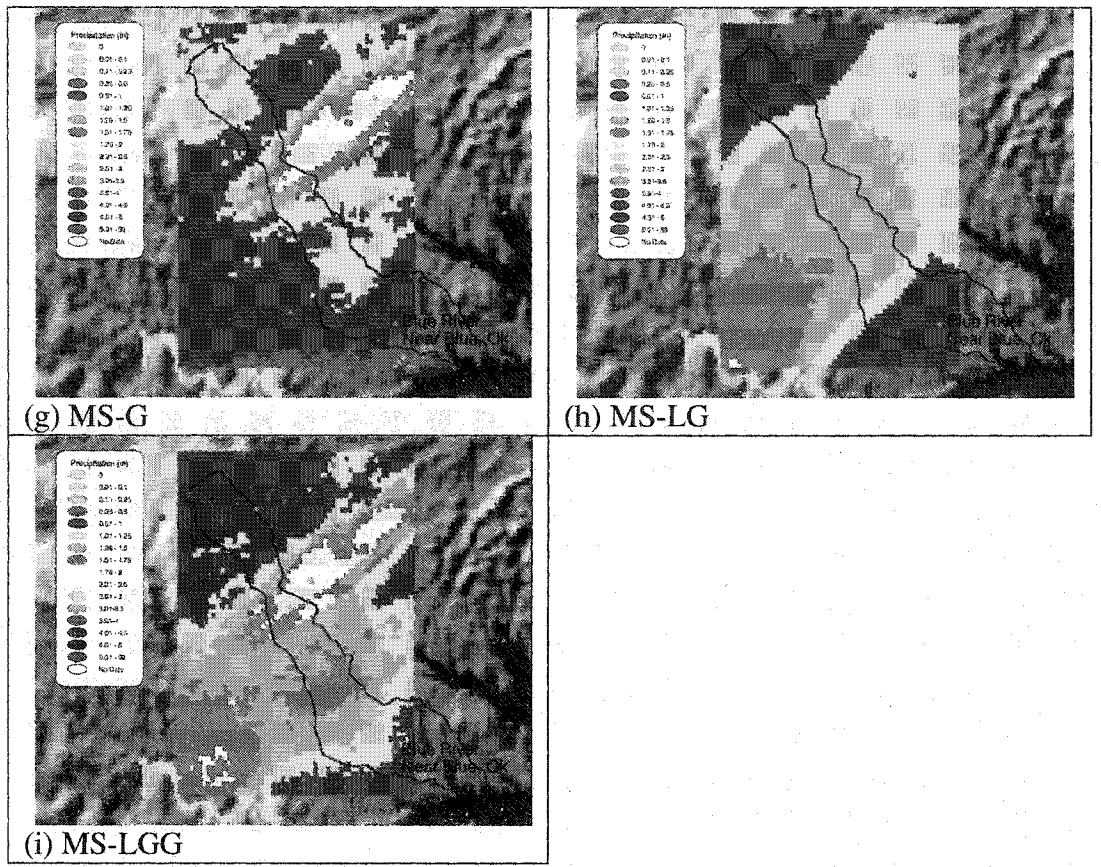


FIG. 4.1. Storm total precipitation plots for the 23 October 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c) radar with mean field bias removed, (d) radar with local bias adjustment, (e) radar with mean field bias removed and local bias adjustment, (f) multisensor, (g) multisensor with mean field bias removed, (h) multisensor with local bias adjustment, and (i) multisensor with mean field bias removed and local bias adjustment. A complete description of each algorithm is provided in Chapter II, section 2.1.4. Values are in inches as indicated in the color bar. Red dots correspond to Mesonet rain gauging locations and cyan triangles are USGS streamgauges,

adjustment technique places a high degree of emphasis on the GAG product (Figs. 4.1a, 4.1d, and 4.1h). This adjustment technique was originally designed for areas that don't have the high density of rain gauge observations associated with the Oklahoma Mesonet. The weighting functions apparently are not optimized for this gauge density, so the GAG field (Fig. 4.1a) essentially overrides the original RAD and MS fields making the GAG, RAD-LG, and MS-LG nearly indistinguishable from one other. The third gauge adjustment strategy appears to be a "middle ground" between the mean field bias and local bias strategies. The radar product with mean field bias removed and local bias adjustment applied (RAD-LGG; Fig. 4.1e) has a similar appearance to the RAD-G product (Fig. 4.1c), but is lighter in accordance with the RAD-LG product (Fig. 4.1d). The MS-LGG field (Fig. 4.1i) also takes on the appearance of being a blend between the MS-G (Fig. 4.1g) and MS-LG product (Fig. 4.1h).

The quantitative component of the hydrologic evaluation begins by creating model parameter ensembles for each input shown in Fig. 4.1. Each rainfall estimator is input to the *Vflo* ™ model for all possible parameter combinations (125). Predicted hydrographs are decomposed into three hydrologic variables: time at which maximum discharge occurred (Time), magnitude of maximum discharge (Peak), and the time-integrated discharge volume normalized by the basin area (Volume). These three variables are also derived from the

observed hydrographs. For reference, the observed hydrograph is shown in Fig. 4.2. The following statistics are used to compare the ensemble mean of predicted to the observed hydrologic variables: *bias*, *MAE*, and *RMSE*. Utilization of the ensemble mean implicitly suggests that the data distribution can be approximated with a Gaussian shape. As will be shown, this may not be the case and undermines the use of parametric statistics. Nonetheless, the simplicity of the calculations is attractive and they reveal fundamental differences between predictions and observations of hydrologic variables. Table 4.1 summarizes the aforementioned statistics compared to observations. Included in the table is the *RPS* value. This score essentially compares the entire pdf of the predicted variables to a single observation. Predictions that are far removed from the observation are penalized more heavily than those falling into nearby categories. The *RPS* is more robust than the parametric statistics in that it considers the entire ensemble of predictions.

Table 4.1 indicates that the Time is predicted most accurately using inputs from the RAD-G rainfall product. All statistics support this conclusion. The mean of the ensemble that uses the MS product as input has the lowest bias, indicating the Time is predicted too early compared to observations. Peak values are predicted accurately using inputs from both RAD and MS-G products. The

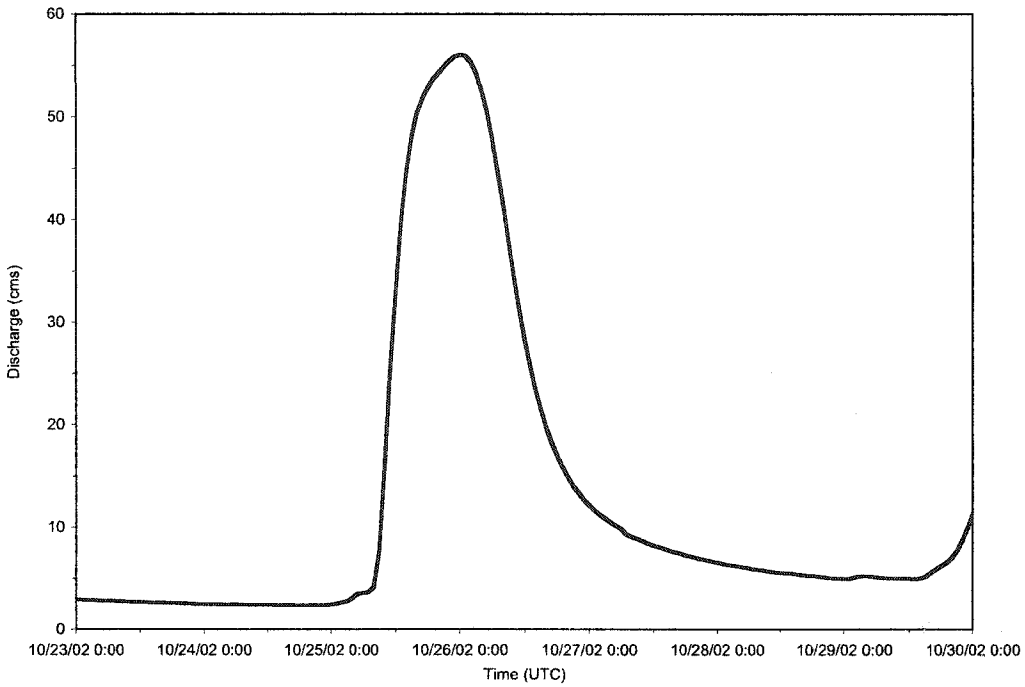


FIG. 4.2. Observed streamflow at the Blue River, near Blue, OK gauging site (USGS #07332500) for the 23 October 2002 case.

TABLE 4.1. Statistical evaluation of hydrologic predictions for the 23 October 2002 case. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). See Chapter II, section 2.1.4 for complete descriptions of rainfall inputs (in first column) and Chapter III, section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.

TIME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.21	0.95	0.39	0.43	1.10
RAD	0.23	0.95	0.39	0.43	1.02
RAD-G	0.18	**0.98**	**0.25**	**0.33**	**0.32**
RAD-LG	0.21	0.95	0.40	0.43	1.14
RAD-LGG	0.16	0.94	0.36	0.44	0.99
MS	0.39	0.93	0.59	0.59	2.14
MS-G	0.22	0.94	0.63	0.63	1.79
MS-LG	0.21	0.95	0.39	0.43	1.10
MS-LGG	0.19	0.94	0.37	0.42	1.04

PEAK	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.21	1.31	23.67	36.65	0.44

RAD	0.23	1.27	**21.39**	31.02	0.43
RAD-G	0.18	1.87	49.17	63.56	1.87
RAD-LG	0.21	1.32	24.08	37.46	0.45
RAD-LGG	0.16	2.17	66.12	88.24	2.42
MS	0.39	0.58	26.45	**29.62**	1.21
MS-G	0.22	**1.15**	21.96	30.39	**0.43**
MS-LG	0.21	1.31	23.86	37.03	0.44
MS-LGG	0.19	1.54	31.52	49.57	0.78

VOLUME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.21	2.03	6.27	8.86	1.86
RAD	0.23	1.96	5.87	7.97	1.88
RAD-G	0.18	2.79	10.89	12.98	4.11
RAD-LG	0.21	2.04	6.33	8.95	1.86
RAD-LGG	0.16	2.84	11.22	13.53	4.14
MS	0.39	**1.15**	**2.31**	**3.42**	**0.51**
MS-G	0.22	1.96	5.87	8.41	1.65
MS-LG	0.21	2.03	6.30	8.93	1.86
MS-LGG	0.19	2.29	7.87	10.47	2.58

TABLE 4.2. Significance levels of RPS differences for the 23 October 2002 case. Significance levels are determined using a resampling test (see Chapter III, section 3.2.2 for complete discussion).

TIME	GAG	RAD	RAD-G	RAD-LG	RAD-LGG	MS	MS-G	MS-LG	MS-LGG
GAG	0.07	0.55	0.99	0.31	0.57	0.99	0.99	0.18	0.40
RAD		0.08	0.99	0.64	0.27	0.99	0.99	0.52	0.19
RAD-G			0.11	0.99	0.99	0.99	0.99	0.99	0.99
RAD-LG				0.07	0.72	0.99	0.99	0.31	0.59
RAD-LGG					0.04	0.99	0.99	0.58	0.38
MS						0.05	0.96	0.99	0.99
MS-G							0.06	0.99	0.99
MS-LG								0.06	0.42
MS-LGG									0.06

PEAK	GAG	RAD	RAD-G	RAD-LG	RAD-LGG	MS	MS-G	MS-LG	MS-LGG
GAG	0.10	0.26	0.99	0.24	0.99	0.99	0.32	0.24	0.98
RAD		0.09	0.99	0.27	0.99	0.99	0.33	0.24	0.98
RAD-G			0.04	0.99	0.97	0.99	0.99	0.99	0.99

RAD-LG			0.08	0.99	0.99	0.29	0.23	0.98
RAD-LGG				0.04	0.99	0.99	0.99	0.99
MS					0.06	0.99	0.99	0.99
MS-G						0.13	0.30	0.99
MS-LG							0.10	0.98
MS-LGG								0.06

VOLUME	GAG	RAD	RAD-G	RAD-LG	RAD-LGG	MS	MS-G	MS-LG	MS-LGG
GAG	0.06	0.13	0.99	0.06	0.99	0.99	0.72	0.06	0.99
RAD		0.05	0.99	0.14	0.99	0.99	0.78	0.14	0.99
RAD-G			0.05	0.99	0.25	0.99	0.99	0.99	0.99
RAD-LG				0.06	0.99	0.99	0.75	0.05	0.99
RAD-LGG					0.04	0.99	0.99	0.99	0.99
MS						0.12	0.99	0.99	0.99
MS-G							0.05	0.74	0.99
MS-LG								0.05	0.99
MS-LGG									0.05

same *RPS* skill is achieved using either product. As indicated by Table 4.2, the statistical significance of the difference in *RPS* values is only 33%, confirming they are essentially the same. The Volume hydrologic variable considers the total flux of water leaving the basin, thus summarizing the prediction in a storm total sense. This variable is the most important in terms of quantifying the hydrologic predictions for watershed management purposes and for evaluating the accuracy of the rainfall inputs. In this case, all statistics indicate the MS product *without any gauge adjustment* is the most accurate input after it has been run through the model using all possible parameter combinations. Mean field bias adjustments to the RAD and MS products result in worse hydrologic predictions. This finding is significant because rain gauges have traditionally been thought of as "ground truth" and a potential calibration source for radar-based algorithms. RPS values for predictions using GAG, RAD-LG, and MS-LG as inputs are nearly identical. Table 4.2 indicates that there is only a 6% chance that these values are statistically different. These statistics confirm the similarity noted in the appearance of the products (Figs. 4.1a, 4.1d, 4.1h).

Data densities are estimated using a Gaussian kernel. In this case, "data" refers to the ensemble of predictions for each variable. The kernel density estimation procedure provides an estimate of the probability distribution function (pdf) describing the ensemble of predicted variables. The word "probability" is used here with the understanding that the model inputs, model structure, and/or

observations may not be perfect. The goal in estimating a pdf here is to benefit from the statistical properties inherent with smooth, differentiable functions. Specifically, the pdfs qualitatively show the data distribution with no dependence on the width or number of categories, a noted problem with histograms. Moreover, the pdfs enable accurate estimation of the 5% and 95% simulation quantiles.

The computed pdfs for Time, Peak, and Volume are shown in Figs. 4.3-4.6. The pdfs supply additional information about the distribution of the hydrologic predictions that is not as apparent in the statistical analysis. For example, Fig. 4.3 reveals the predictions have a distinct bimodal shape. This double peak indicates the parameter settings used in the *Vflo* ™ model produce two different modes of behavior for all model inputs. All products except RAD-G and RAD-LGG have a higher density at earlier Times, thus favoring the first mode of behavior. Figure 4.4 also reveals a bimodal behavior, especially with the pdfs that use inputs resulting in relatively lower Peak predictions (e.g., MS). The secondary peaks in the RAD-G and RAD-LGG are less distinct and result in long tails at higher Peak values. The secondary peaks in the pdfs are more evident with Volume predictions (Fig. 4.5). The MS pdf appears to be significantly different from the other pdfs. As it turns out, the statistical levels of the *RPS* differences between the MS input and all other inputs exceed 99% (see Table

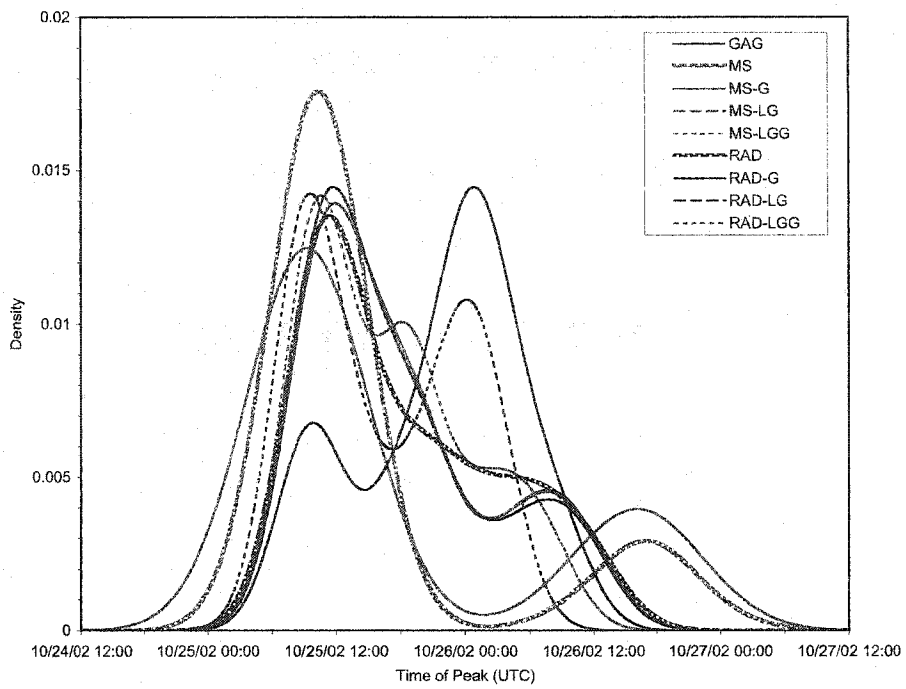


FIG. 4.3. Probability density functions of the predicted time of maximum discharge (Time) for the 23 October 2002 case. The color coding of the curves (noted in the legend) corresponds to the different model inputs used to construct the model parameter ensembles. Observed Time is 0000 UTC 26 October 2002.

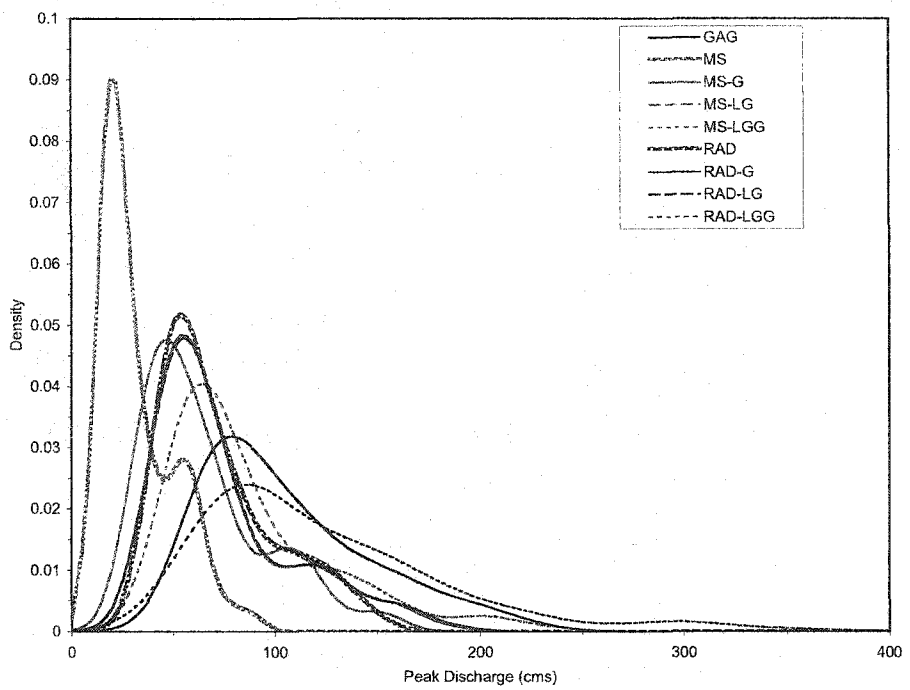


FIG. 4.4. As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak). Observed Peak is 56.7 cubic meters per second (cms).

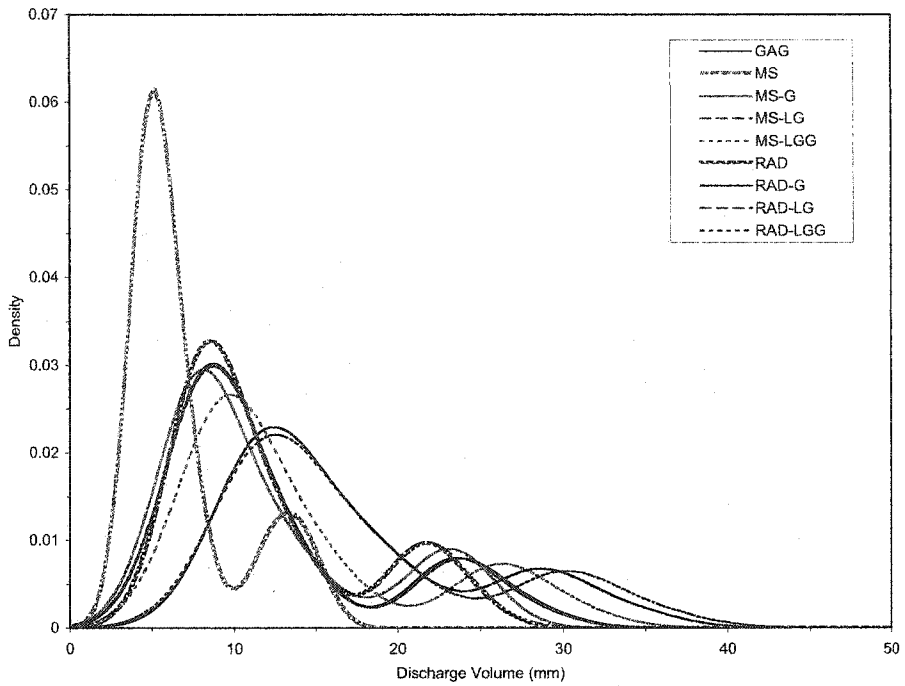


FIG. 4.5. As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area. Observed Volume is 6.1 mm.

4.3), thus providing further support of this inference. The pdfs for all predicted variables (Figs. 4.3-4.5) show that the GAG, RAD-LG, and MS-LG inputs are nearly identical. This observation confirms the findings in the subjective comparison of storm total plots and the statistical analysis. In this case, these three products are equivalent, visually and statistically. Their use in a model input ensemble or combined input-parameter ensemble may not be warranted as these inputs do not result in different hydrologic outcomes. In any case, it is informative to the developers of quantitative precipitation algorithms to know that these products are the same over the Blue River Basin.

The final component of the hydrologic evaluation for this case involves deriving the region in the phase space where 90% of the simulations fall. This is accomplished by iteratively summing each pdf in Figs. 4.3-4.5 so that they become cumulative distribution functions (cdfs; not shown). The Time, Peak, and Volume values corresponding to cumulative densities of .05 and .95 are referred to as the 5% and 95% simulation bounds. 90% of the simulations fall within those bounds. The distance between the 5% and 95% quantile is an approximate measure of the ensemble spread. A wide spread suggests a high degree of uncertainty. A smaller spread implies more confidence, yet it is imperative that the spread encompasses the observed variable. If not, then a high confidence will be placed on an incorrect prediction. A nonparametric measure of central tendency is obtained by finding the value corresponding to a cumulative density

of .50. This is the median of the simulations. Plots of the 5%, 50%, and 95% quantiles are shown in Figs. 4.6-4.8. The observed hydrologic variables are shown in these plots as a double line. This convention is a reminder that observations of streamflow have their own uncertainty, but have been found to be only 3-6%.

Figure 4.6 shows that all ensembles envelop the observed Time regardless of the rainfall product being used as input. Wider simulation bounds are evident with the MS and MS-G inputs suggestive of more uncertainty. The ensemble medians for the timing of peak discharge all occur slightly earlier than observed. The central tendency of the RAD-G input indicates this product is capable of producing the most accurate Time predictions for this case. Statistics in Table 4.1 support this latter conclusion.

The capabilities of each ensemble to predict the peak discharge are revealed in Fig. 4.7. Similar to Fig. 4.6, the simulation bounds from all ensembles encompass the observed Peak. Use of the MS product as the model input results in the smallest spread of Peak predictions. In this case, the central tendency of the predictions is less than the observed. The apparent underforecast is indicated explicitly in Table 4.1 with the *bias* statistic that uses the ensemble mean. The medians of the GAG, RAD, RAD-LG, and MS-LG ensembles lie very near the observed value. The spread of the ensembles are different, however, which results in variable *RPS* values. In fact, the lower uncertainty with the MS-

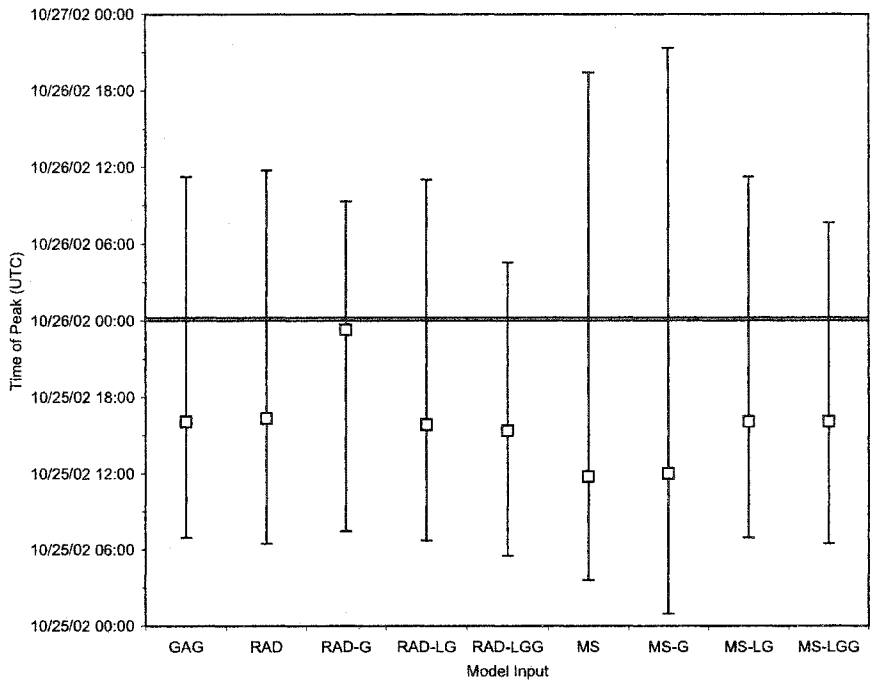


FIG. 4.6. Simulation bounds of the predicted time of maximum discharge (Time) for the 23 October 2002 case. The open boxes refer to the 50% quantile (median), while the bars correspond to the 5% and 95% quantiles. The horizontal double line is the observed Time.

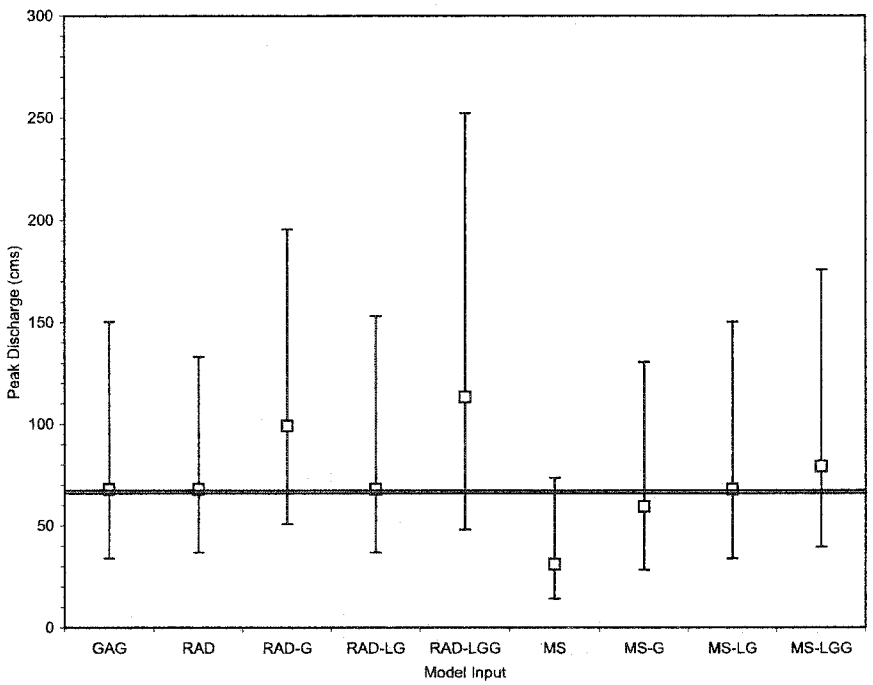


FIG. 4.7. As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 23 October 2002 case.

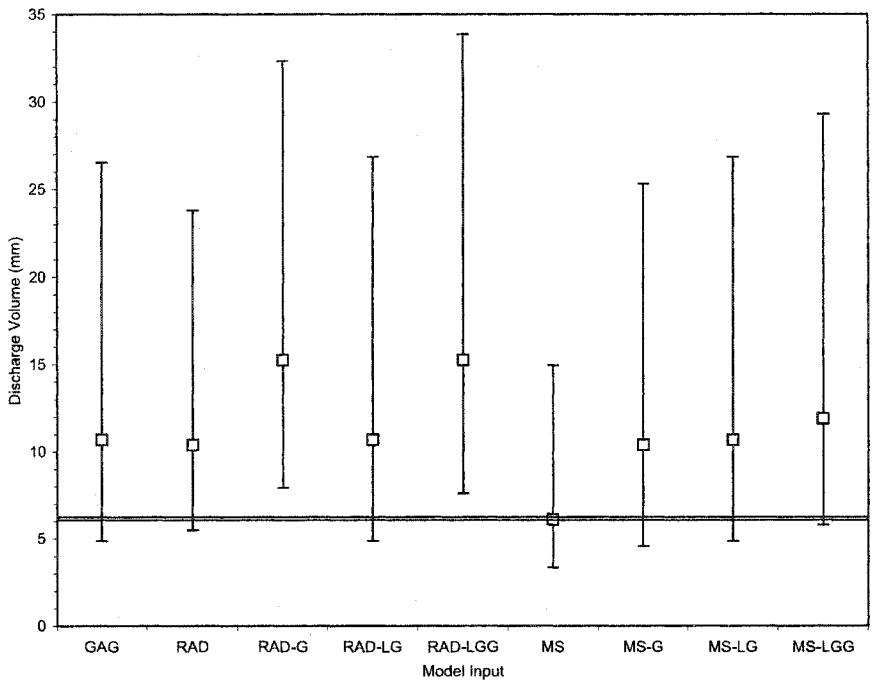


FIG. 4.8. As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 23 October 2002 case.

G ensemble, despite the underforecast noted from the ensemble median, produces a skill score equivalent to the ensemble that uses RAD model inputs.

Figure 4.8 shows Volume predictions using inputs from the RAD-G and RAD-LGG rainfall products do not give the observed system behavior. The rainfall amounts from these two algorithms are too high which results in 90% of the Volume predictions to be overforecast. The median of the total discharge volume predictions using MS inputs agrees rather well with the observed Volume. In addition, the uncertainty bounds are relatively small with this rainfall input. A better *RPS* skill score results. The hydrologic evaluation for the 23 October 2002 indicates the MS algorithm as an input results in the most accurate hydrologic predictions for Volume. It should be noted, however, that this initial conclusion is conditioned on perfect model physics and observations of streamflow. Errors in these components of the modeling process can be structured to favor a given QPE input that may be biased. Identification of possible model structural uncertainties is explored in section 4.2.2 of this chapter. Moreover, additional cases are examined below to determine the rainfall algorithm that produces the most accurate inputs from the hydrologic modeling perspective.

4.1.2. 28 October 2002 Case

The next case used for QPE evaluation yielded approximately 40% of the Volume than from the 23 October 2002 case discussed above (refer to Table 3.1).

Rainfall products from the RAD, RAD-G, RAD-LG, and RAD-LGG are not available for this hydrologic evaluation due to a failure in the archiving process. The overall skill of the available algorithms is determined as well as the accuracy of different gauge adjustment strategies. The storm total precipitation amounts are shown in Figs. 4.9a-e. Subjective comparison between the products reveals much more smoothness with the GAG (Fig. 4.9a), MS-LG (Fig. 4.9d), and MS-LGG (Fig. 4.9e) products. These rainfall estimates rely heavily on point rain gauge amounts that have been objectively interpolated to grid points between the gauges. The analysis scheme tends to produce light rainfall in lower reaches of the basin. The MS product (Fig. 4.9b), on the other hand, utilizes information from radars and satellite. Individual storm cells that have moved from a southwesterly to northeasterly direction are resolved with this product. Moreover, the precipitation appears to fall in upper reaches of the basin, while locations downstream receive no rainfall at all. Mean field bias adjustments to the MS product (MS-G; Fig. 4.9c) result in a spatial pattern similar to the MS field, but the amounts are increased.

The observed hydrograph is shown in Fig. 4.10. Derived Time, Peak, and Volume variables for observed and predicted hydrographs are used in a quantitative analysis of the ensembles' performance (Table 4.3). The Time is predicted most accurately using either the MS-G or MS-LGG rainfall estimates as inputs to the model parameter ensembles. The *RPS* score is improved slightly

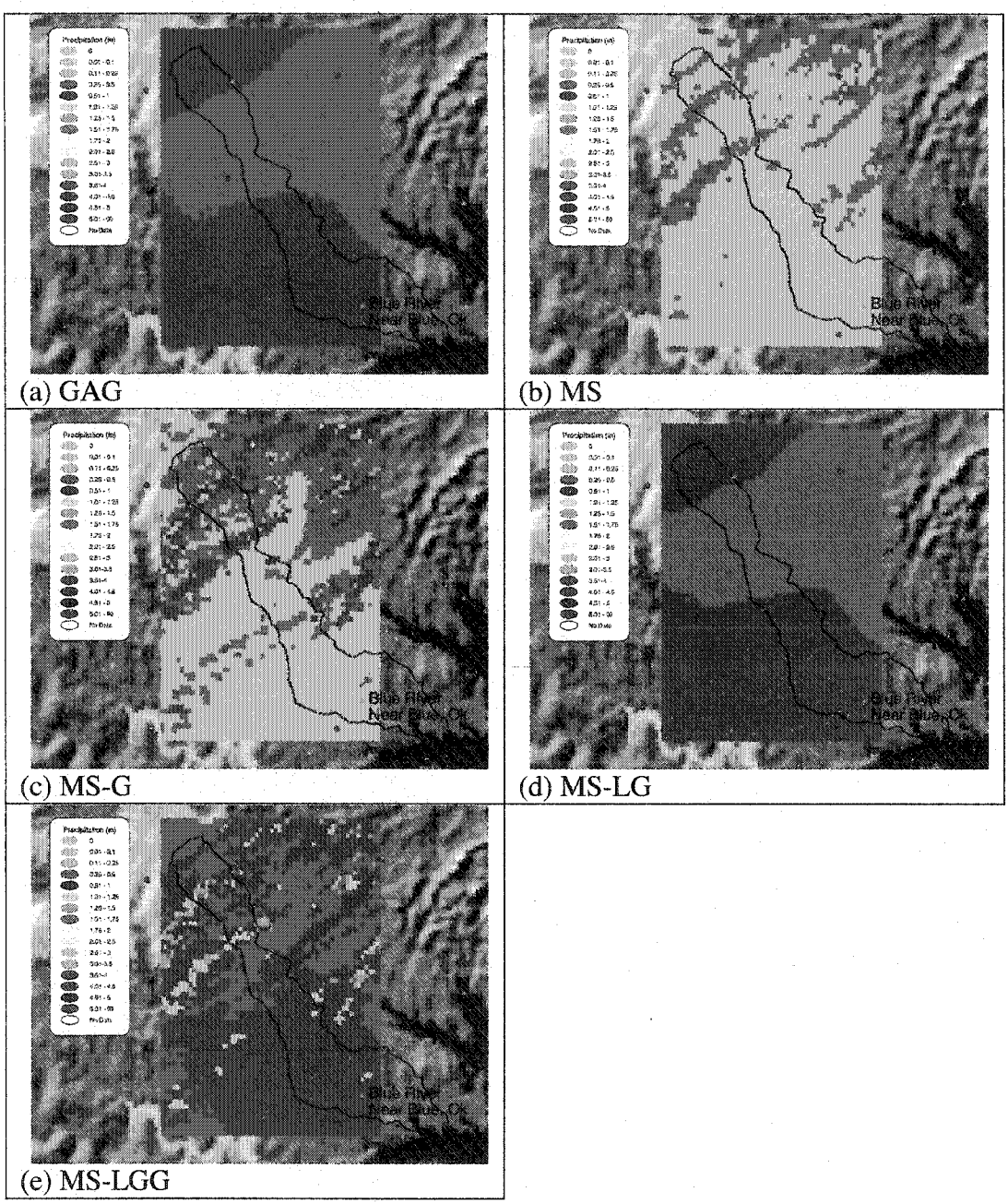


FIG. 4.9. As in Fig. 4.1 but for storm total precipitation plots for the 28 October 2002 case from the QPESUMS products: (a) gauge-only, (b) multisensor, (c) multisensor with mean field bias removed, (d) multisensor with local bias

adjustment, and (e) multisensor with mean field bias removed and local bias adjustment.

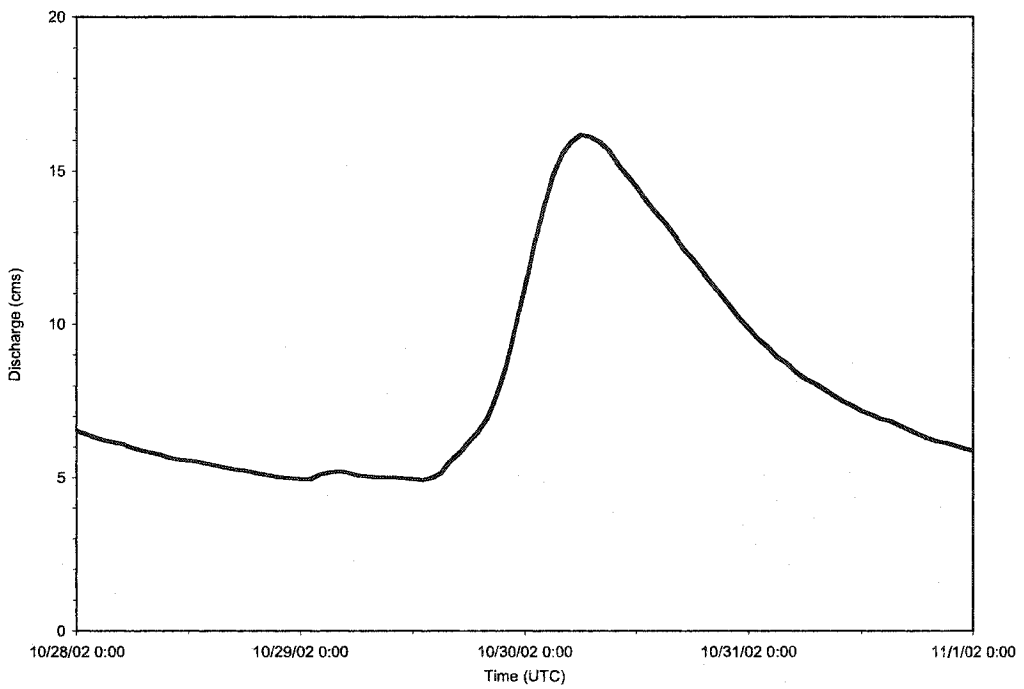


FIG. 4.10. As in Fig. 4.2 but for observed hydrograph for 28 October 2002 case.

TABLE 4.3. As in Table 4.1 but for statistical evaluation of hydrologic predictions for the 28 October 2002 case.

TIME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	0.88	0.69	0.69	3.05
MS	0.74	1.25	1.26	1.31	3.65
MS-G	0.42	1.12	**0.58**	**0.61**	2.13
MS-LG	0.14	0.88	0.69	0.69	3.05
MS-LGG	0.13	**0.91**	0.64	0.67	**2.11**

PEAK	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	2.99	32.64	40.20	4.28
MS	0.74	**0.63**	**8.10**	**8.74**	1.25
MS-G	0.42	1.43	9.10	13.73	**0.69**
MS-LG	0.14	3.00	32.74	40.30	4.30
MS-LGG	0.13	3.30	37.69	44.62	4.88

VOLUME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	3.27	5.23	6.25	5.14
MS	0.74	**1.31**	**0.71**	**0.95**	**0.70**
MS-G	0.42	1.79	1.82	2.14	2.19

MS-LG	0.14	3.19	5.04	6.09	4.82
MS-LGG	0.13	3.75	6.32	7.22	5.97

TABLE 4.4. As in Table 4.2 but for significance levels of RPS differences for the 28 October 2002 case.

TIME	GAG	MS	MS-G	MS-LG	MS-LGG
GAG	0.05	0.99	0.99	0.04	0.99
MS		0.04	0.99	0.99	0.99
MS-G			0.08	0.99	0.42
MS-LG				0.05	0.99
MS-LGG					0.05

PEAK	GAG	MS	MS-G	MS-LG	MS-LGG
GAG	0.05	0.99	0.99	0.17	0.99
MS		0.07	0.99	0.99	0.99
MS-G			0.07	0.99	0.99
MS-LG				0.05	0.99
MS-LGG					0.05

VOLUME	GAG	MS	MS-G	MS-LG	MS-LGG
GAG	0.08	0.99	0.99	0.99	0.99
MS		0.10	0.99	0.99	0.99
MS-G			0.06	0.99	0.99
MS-LG				0.08	0.99
MS-LGG					0.23

with the MS-LGG input, but the statistical significance of the *RPS* differences is only 42% as indicated in Table 4.4. Conclusions regarding the Peak predictions can differ when considering the mean of the ensemble predictions versus the entire pdf. The MS input has the best *bias*, *MAE*, and *RMSE*, while the MS-G algorithm produces the best *RPS* score when the entire ensemble of Peak predictions is considered. The confidence of the *RPS* differences between the MS and MS-G inputs is 99% (Table 4.4). Volume predictions are most skillful using MS estimates as an input to the model parameter ensemble. Statistics describing the ensemble mean and the entire pdf support this conclusion, while the products that rely more heavily on rain gauge estimates (GAG, MS-LG, and MS-LGG) result in lower *RPS* skill scores. The *bias*es using the ensemble means from these products exceed a factor of 3. It is likely that the point rain gauge estimates for this rainfall event do not accurately represent the spatial variable of the true rainfall that fell over the basin.

The derived pdfs from the ensembles reveal additional information about the hydrologic predictions produced with the different rainfall inputs. Figure 4.11 shows the pdfs of the Time predictions. Similar to the analysis of pdfs in section 4.1.1, two distinct modes of behavior are noted in the predictions. Predictions from the GAG, MS-LG, and MS-LGG algorithms all favor the earlier mode. The MS-G pdf is bimodal and nearly symmetric. Apparently, the model parameter ensemble that uses MS-G inputs straddles the inflection point where the different modes of behavior reside. The Time predictions from the MS ensemble favor the second, later mode. Figure 4.12 shows the pdfs from the differing rainfall inputs also exhibit bimodal shapes. In this case, higher densities are associated with smaller Peak predictions for all inputs. Predictions from the MS algorithm are the lowest which is also the case with the 23 October 2002 event. The shapes of pdfs help explain the discrepancy in the statistics that rely on the ensemble mean versus the entire pdf (Table 4.3). The mean of the Peak predictions from the MS ensemble is sensitive to the secondary relative maximum. This results in an expected value that is greater than the median. Mean field bias adjustments increase the magnitude of the MS Peak predictions resulting in better agreement with observations as indicated by the low *RPS* value (Table 4.3). The pdfs for predicted Volume for this case are shown in Fig. 4.13. The MS algorithm produces the lowest Volume predictions, and a bimodal shape is evident once again. The mean field bias adjusted MS product results in higher Volume

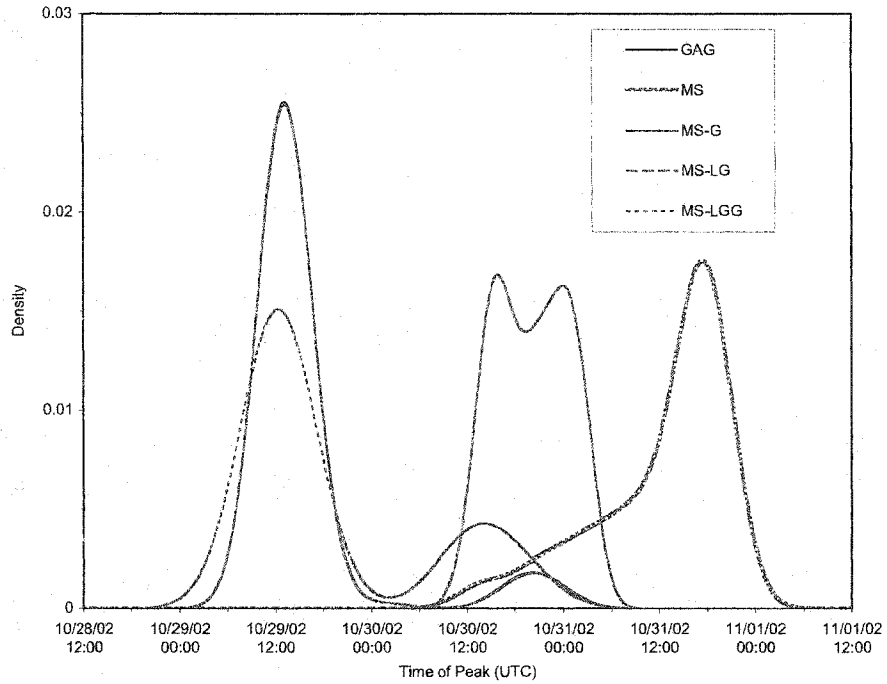


FIG. 4.11. As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 28 October 2002 case. Observed Time is 0600 UTC 30 October 2002.

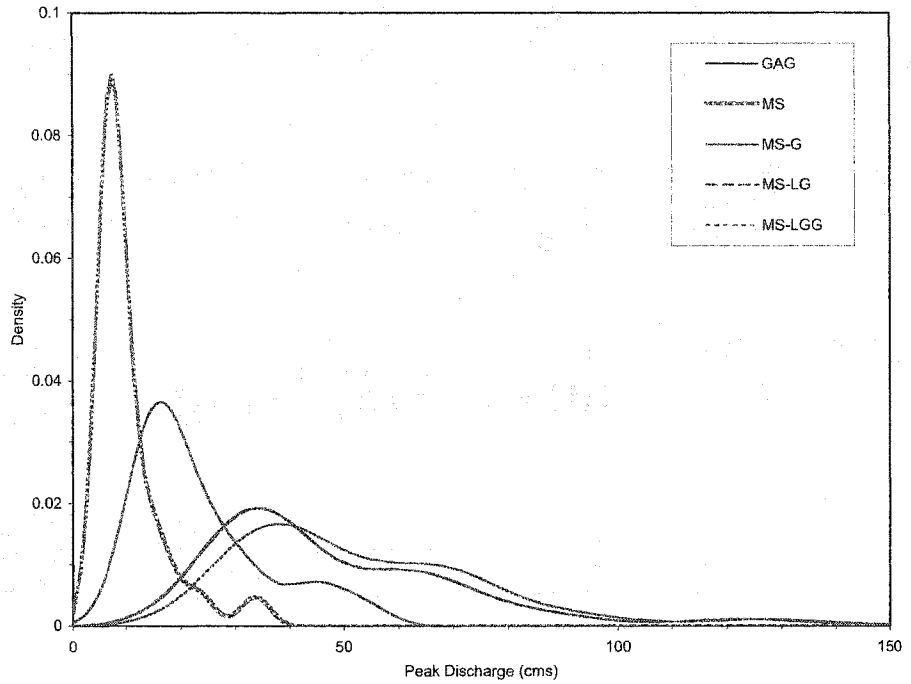


FIG. 4.12. As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case. Observed Peak is 16.4 cubic meters per second (cms).

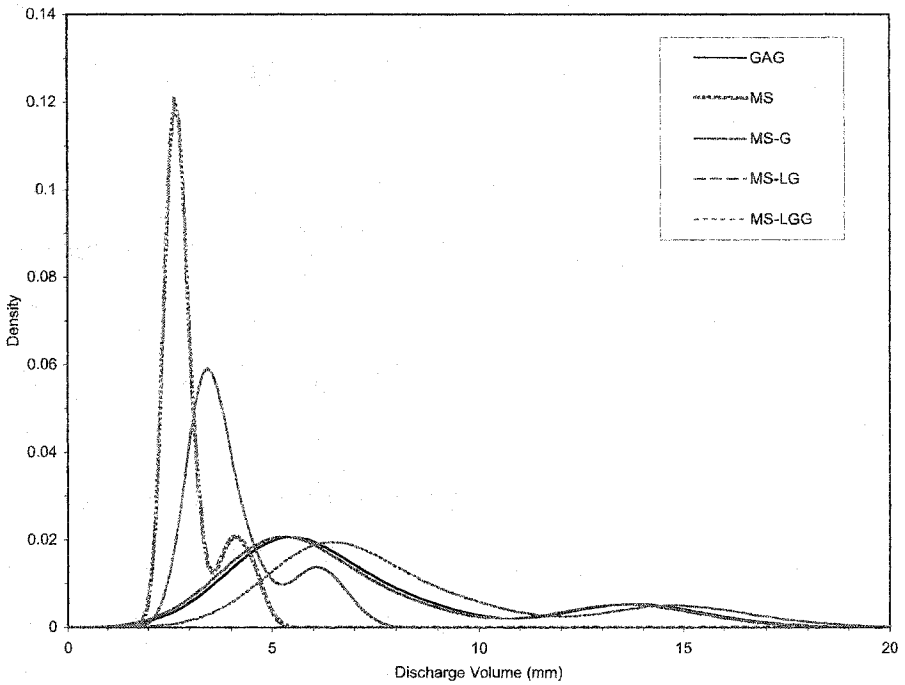


FIG. 4.13. As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 28 October 2002 case. Observed Volume is 2.3 mm.

predictions, but not as large as those produced by the GAG, MS-LG, or MS-LGG algorithms. The pdfs shown in Figs. 4.11-4.13 reveal the shapes of the distributions for the predicted hydrologic variables. They are also used to derive simulation quantiles and spreads providing metrics for accuracy and uncertainty.

Figure 4.14 shows the 90% simulation bounds and predicted median for timing of peak discharge using different rainfall algorithms as model inputs. The spread is a result of parametric uncertainty. The pdf from the MS ensemble (Fig. 4.11) indicates that the predictions tend toward the second mode of behavior. In this case, relative maxima in the pdfs that occur at earlier times are evidently more correct than the later ones. Rainfall inputs that rely heavily on rain gauge data (e.g., GAG, MS-LG, MS-LGG) produce simulations that encompass the observed time to peak. Ensembles of Time predictions using MS and MS-G inputs do not accurately provide the observed behavior for this case.

The width of the simulation bounds for ensembles using GAG, MS-LG, and MS-LGG inputs is relatively wide for Peak predictions (Fig. 4.15). More importantly, the 90% simulation limits do not include the observed peak discharge for the aforementioned ensembles. Peak predictions using either MS or MS-G inputs result in better agreement with observations. The MS-G ensemble has a median simulation that aligns very near the observed peak discharge. The MS ensemble has the narrowest uncertainty bounds which include the observed value.

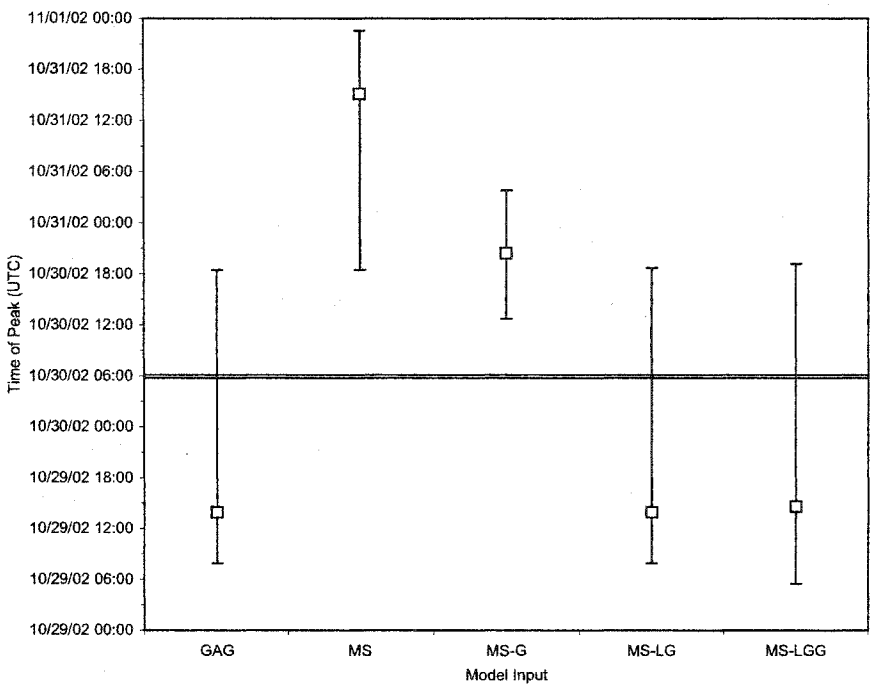


FIG. 4.14. As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 28 October 2002 case.

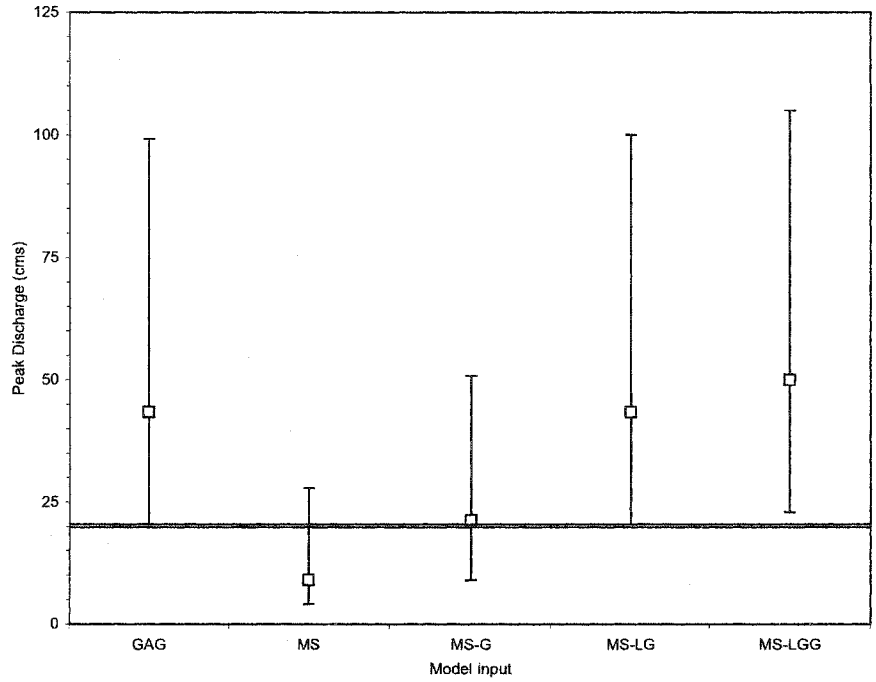


FIG. 4.15. As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case.

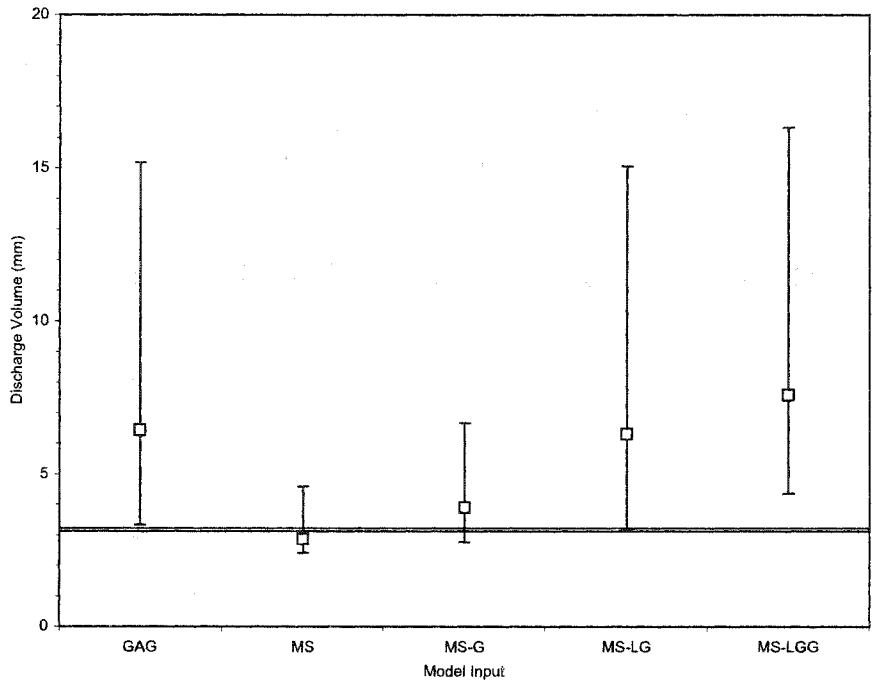


FIG. 4.16. As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 28 October 2002 case.

Table 4.3, however, indicates that the most skillful Peak predictions are associated with the model parameter ensemble that uses MS-G inputs.

The simulation bounds for the Volume hydrologic variable are similar to Peak predictions for this case (Figs. 4.15-4.16). MS and MS-G inputs yield an ensemble of Volume predictions that cluster near the observed Volume well. The MS ensemble has very narrow uncertainty bounds, and the central tendency of the ensemble of predictions is close to observations. From Tables 4.3-4.4, adjustment procedures that rely heavily on individual gauge amounts do not result in more accurate Volume predictions from the hydrologic model. Figures 4.9a, 4.9d, and 4.9e indicate that the techniques used to adjust QPEs based on point data result in smooth, interpolated accumulations, especially in lower reaches of the Blue River Basin. Note that there are no rain gauge measurements collected by the Mesonet in the Blue River Basin. MS accumulations on the other hand utilize radar and satellite data to determine the spatial character of the rainfall field. Products that rely more heavily on these remote sensing systems resolve no rainfall at all near the basin outlet. Moreover, notice how rainfall maxima in the MS field (narrow part of basin; Fig. 4.9b) occur in regions where minima occur in the GAG and MS-LG accumulation products (Figs. 4.9a and 4.9d). The MS and gauge-based products are proven to be significantly different in the statistical analysis. Results show that model parameter ensembles using the MS rainfall algorithm as input provide the best predictions for the Volume variable. The MS rainfall input is

believed to be the least biased and most accurate input based on this hydrologic evaluation. It is recognized, however, that the time of peak discharge is not well forecast using the MS inputs. Two modes of behavior (i.e., two relative maxima in the pdfs) have been noted in nearly all of the ensembles' pdfs for Time, Peak, and Volume. The MS ensemble favors the later, less accurate region of the Time phase space. Bimodal pdfs are also noted for the Peak and Volume predictions with all of the ensembles tending toward the first, lower mode of behavior. Explanations for these parametric or possibly model structural modes are provided in section 4.1.4.

4.1.3. 03 December 2002 Case

The final case used for the hydrologic evaluation component of this study produced a slightly larger Volume but smaller peak discharge than the 28 October 2002 case. The storm total accumulations are shown in Figs. 4.17a-e. Multisensor products are not available for this case due to a failure in the archiving procedure. The observed hydrograph is shown in Fig. 4.18. There is agreement in the amounts and locations of precipitation over the Blue River Basin between the independent GAG (Fig. 4.17a) and RAD (Fig. 4.17b) products. The radar product with mean field bias adjustment (RAD-G; Fig. 4.17c), however, indicates that overestimation is prevalent over a larger, statewide domain requiring reductions in the QPEs. Both RAD (Fig. 4.17b) and RAD-G (Fig.

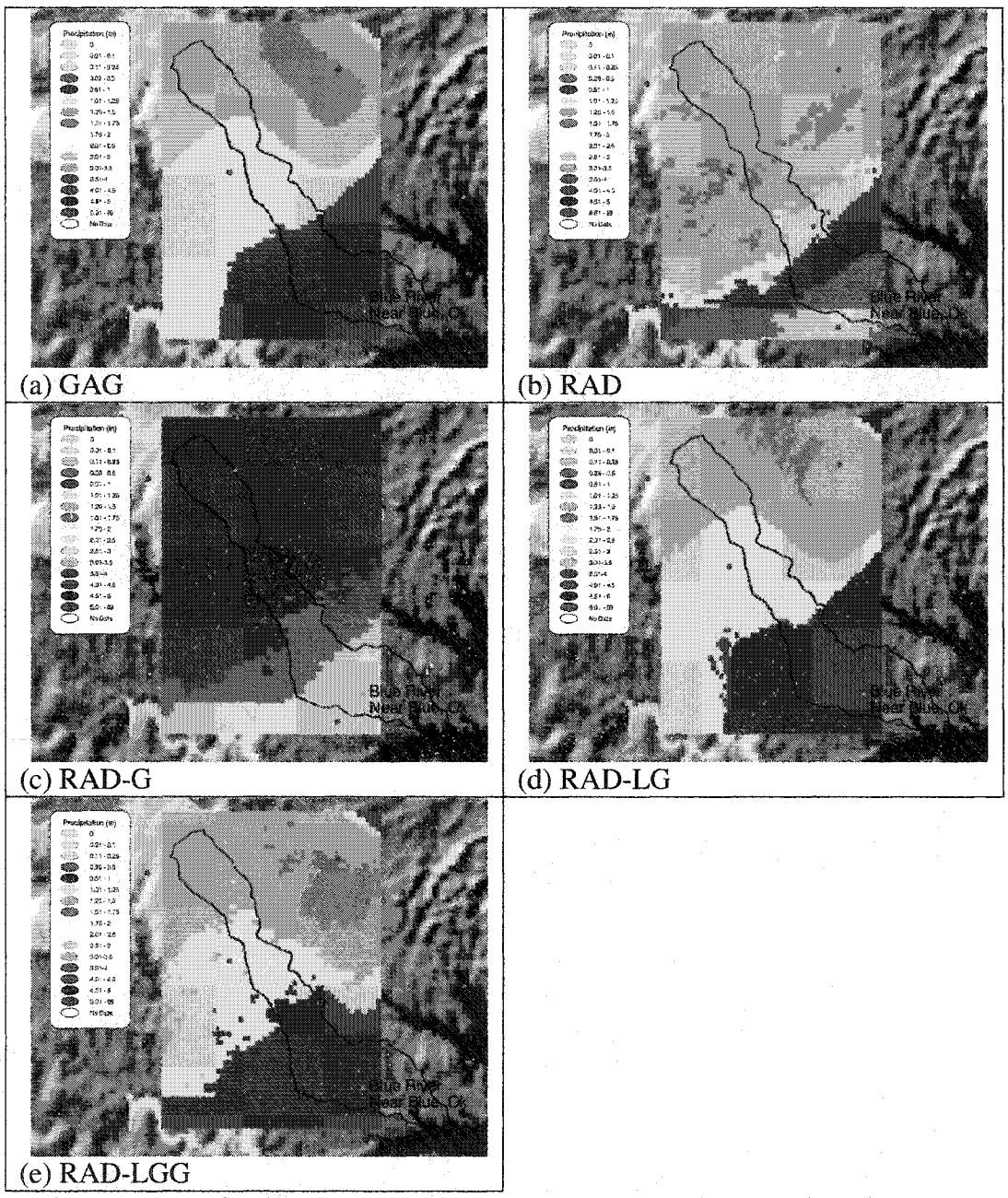


FIG. 4.17. As in Fig. 4.1 but for storm total precipitation plots for the 03 December 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c)

radar with mean field bias removed, (d) radar with local bias adjustment, and (e) radar with mean field bias removed and local bias adjustment.

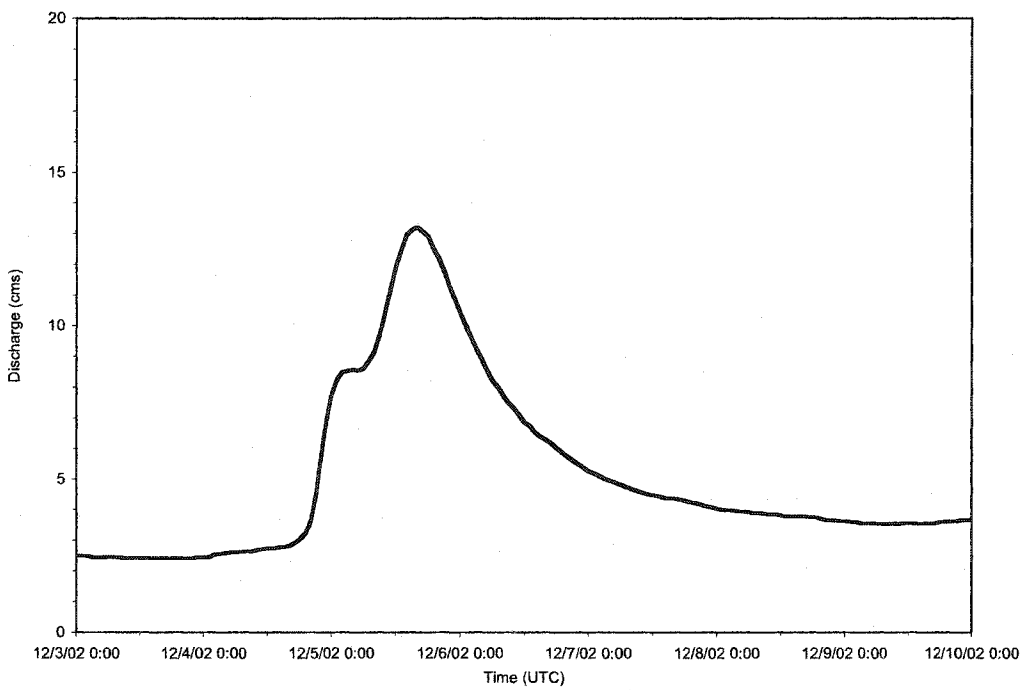


FIG. 4.18. As in Fig. 4.2 but for observed hydrograph for 03 December 2002 case.

4.17c) products show a linear discontinuity near the bottom of the images. This artifact corresponds to the equidistant point between the Oklahoma City, OK and Fort Worth, TX WSR-88D radars (KTLX and KFWS). QPE SUMS mosaics the data from the adjacent radars by choosing the beam that measures reflectivity closest to the surface. At equidistant points, both radars should measure equivalent reflectivity values. However, these images indicate that KTLX estimates much higher rainfall amounts. A recent study by Gourley et al. (2003) shows that the KTLX radar reports relatively high reflectivity values relative to its neighboring radars. This apparent miscalibration of the radar hardware may be the root cause requiring the mean field bias adjustment to lower QPEs in the RAD-G product (Fig. 4.17c) significantly. The quantitative analysis provided herein will illuminate this potential error source in radar-based QPE.

Tables 4.5-4.6 show a statistical evaluation of the different ensembles for all three derived variables. In this case, statistics based on the ensemble mean agree with the *RPS* values based on the entire data distribution. All statistics indicate Time is predicted best using RAD rainfall as inputs for model parameter ensembles, while the ensemble using RAD-G inputs results in the worst Time predictions. Different if not opposite conclusions are drawn for Peak and Volume predictions though. There are few distinguishing characteristics between the GAG, RAD, RAD-LG, and RAD-LGG ensembles (Table 4.5), in agreement with observations of storm total precipitation maps (Figs. 4.17a, 4.17b, 4.17d, and

TABLE 4.5. As in Table 4.1 but for statistical evaluation of hydrologic predictions for the 03 December 2002 case.

TIME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	0.88	0.69	0.69	3.05
MS	0.74	1.25	1.26	1.31	3.65
MS-G	0.42	1.12	**0.58**	**0.61**	2.13
MS-LG	0.14	0.88	0.69	0.69	3.05
MS-LGG	0.13	**0.91**	0.64	0.67	**2.11**

PEAK	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	2.99	32.64	40.20	4.28
MS	0.74	**0.63**	**8.10**	**8.74**	1.25
MS-G	0.42	1.43	9.10	13.73	**0.69**
MS-LG	0.14	3.00	32.74	40.30	4.30
MS-LGG	0.13	3.30	37.69	44.62	4.88

VOLUME	RO Coeff	Bias	MAE	RMSE	RPS
GAG	0.14	3.27	5.23	6.25	5.14
MS	0.74	**1.31**	**0.71**	**0.95**	**0.70**
MS-G	0.42	1.79	1.82	2.14	2.19

MS-LG	0.14	3.19	5.04	6.09	4.82
MS-LGG	0.13	3.75	6.32	7.22	5.97

TABLE 4.6. As in Table 4.2 but for significance levels of RPS differences for the 03 December 2002 case.

TIME	GAG	RAD	RAD-G	RAD-LG	RAD-LGG
GAG	0.06	0.90	0.99	0.14	0.31
RAD		0.05	0.99	0.86	0.83
RAD-G			0.04	0.99	0.99
RAD-LG				0.05	0.34
RAD-LGG					0.05

PEAK	GAG	RAD	RAD-G	RAD-LG	RAD-LGG
GAG	0.08	0.99	0.99	0.08	0.74
RAD		0.06	0.99	0.99	0.95
RAD-G			0.04	0.99	0.99
RAD-LG				0.09	0.75
RAD-LGG					0.06

VOLUME	GAG	RAD	RAD-G	RAD-LG	RAD-LGG

GAG	0.12	0.63	0.99	0.95	0.94
RAD		0.13	0.99	0.96	0.96
RAD-G			0.06	0.99	0.99
RAD-LG				0.10	0.09
RAD-LGG					0.11

4.17e). The ensemble of RAD-G Peak predictions is much lower compared to others as indicated by the ensemble mean *bias* statistic. The *RPS* skill value is also best with the RAD-G product. Similar results are obtained with Volume predictions. The RAD-G simulations are much less biased and more skillful. Evidently, lower amounts estimated by the RAD-G rainfall algorithm are more in tune with an ensemble of predicted hydrographs from the *Vflo* ™ model. However, the timing of peak discharge is erroneously reduced with the RAD-G input.

Figures 4.19-4.21 reveal pdfs of predicted Time, Peak, and Volume for ensembles using different QPE inputs. The data distributions are a result of model parametric uncertainty. As noted in previous evaluation cases, a bimodal shape to the pdfs is evident with every ensemble. The Time pdfs (Fig. 4.19) show the most extreme bimodal behavior. The RAD-G product favors the first mode of behavior. All other ensemble pdfs are bimodal, but have maximum densities at later times. In this case, the secondary bumps in the pdfs agree with observed

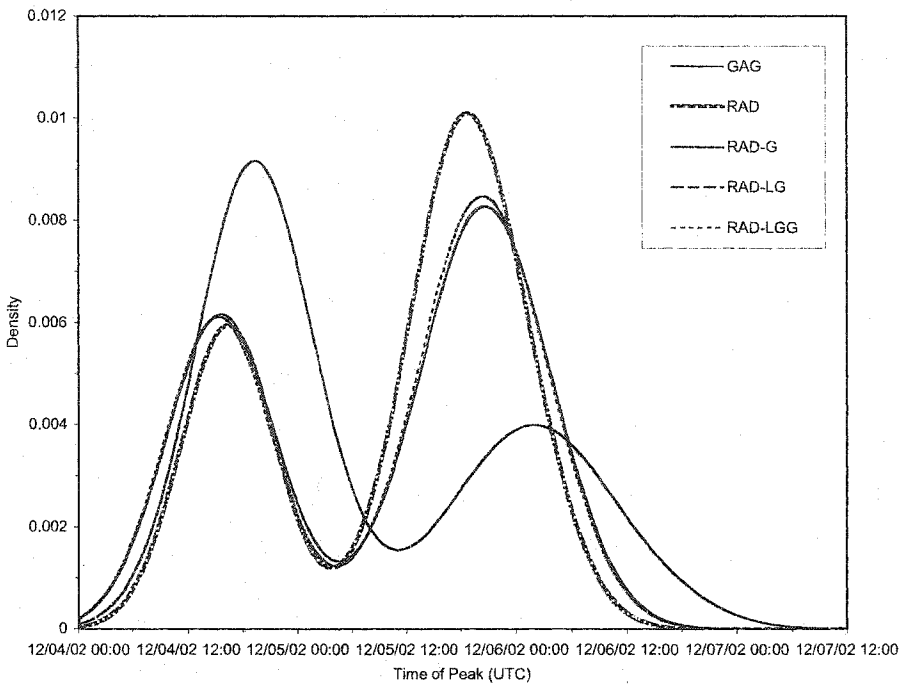


FIG. 4.19. As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 03 December 2002 case. Observed Time is 1600 UTC 05 December 2002.

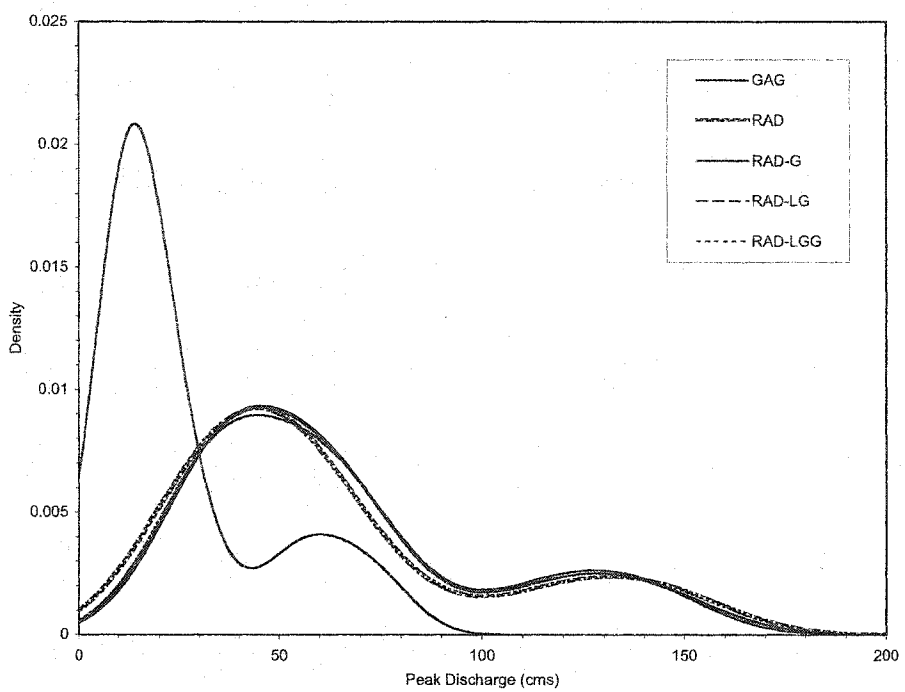


FIG. 4.20. As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case. Observed Peak is 13.4 cubic meters per second (cms).

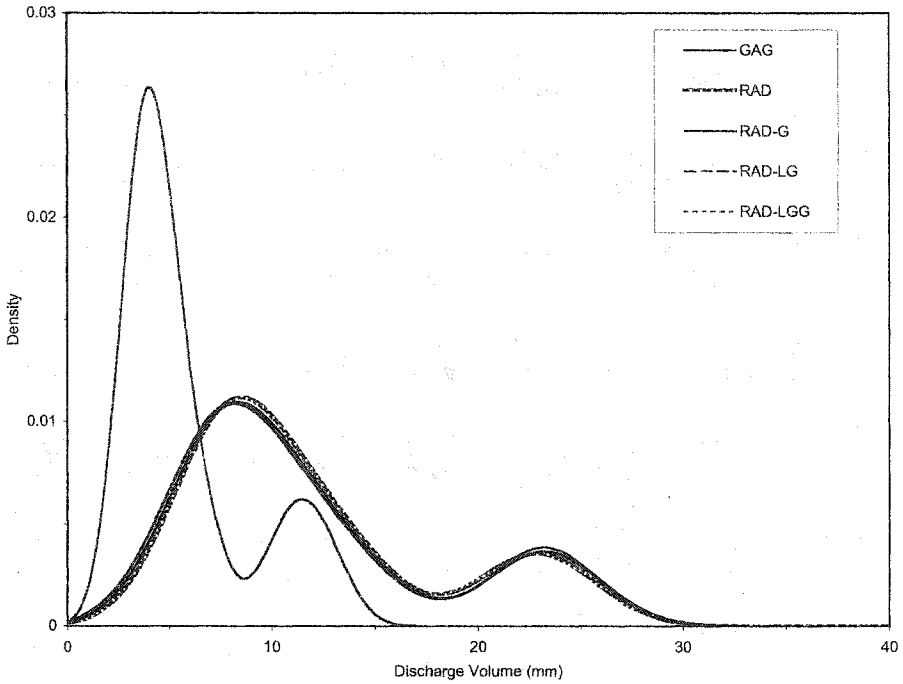


FIG. 4.21. As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 03 December 2002 case. Observed Volume is 2.5 mm.

time to peak more closely. Bimodal shapes are noted for all Peak pdfs in Fig. 4.20 with higher densities occurring with relatively lower Peaks. As mentioned in the statistical comparison, there are very few distinguishing characteristics between the GAG, RAD, RAD-LG, and RAD-LGG pdfs. The Peak pdfs (Fig. 4.20) indicate the RAD-G product is an outlier. Very similar results are obtained for predictions of Volume (Fig. 4.21). All predicted pdfs are bimodal and favor the first mode of behavior. The RAD-G product produces much lower Volume predictions, while the other estimates result in Volume pdfs that are nearly identical.

The uncertainty associated with the parameter ensembles is shown in Fig. 4.22-4.24. The width of the uncertainty bounds for Time predictions, regardless of the model input, spans over approximately two days. The high degree of uncertainty with these predictions is a result of the nearly symmetrical, bimodal shape to the pdfs (Fig. 4.19). The simulation median for the RAD-G ensemble is shown to favor the earlier mode of behavior, while the other medians lie closer to the observed, later Time. All inputs are capable of producing realistic, behavioral predictions in terms of predicting the time of maximum discharge. However, the RAD-G product has a tendency to predict earlier than observed Time variables for this case.

The 90% simulation bounds for all ensembles encompass the observed peak discharge value (Fig. 4.23). As indicated in previous analyses of this case,

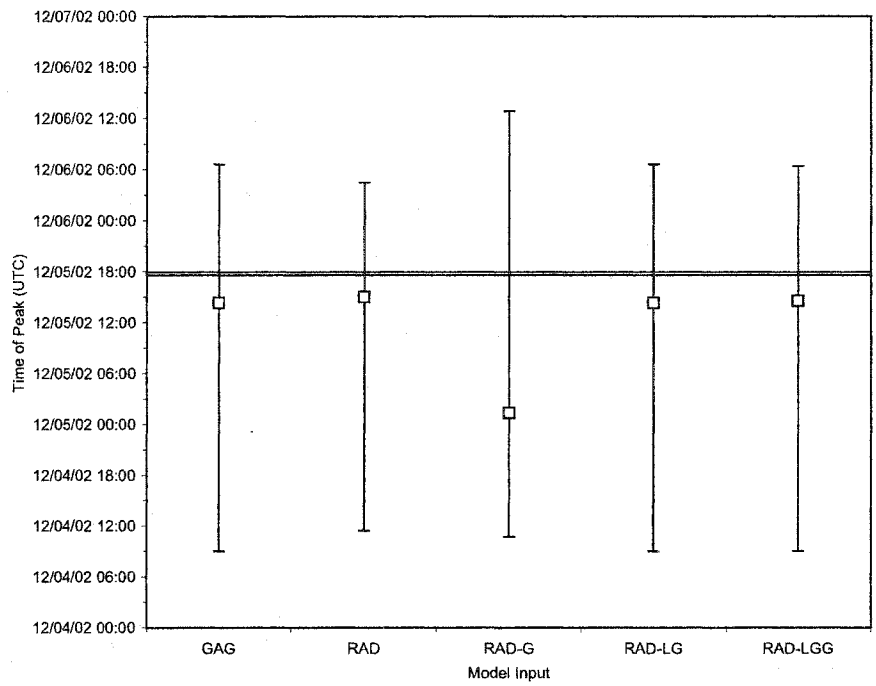


FIG. 4.22. As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 03 December 2002 case.

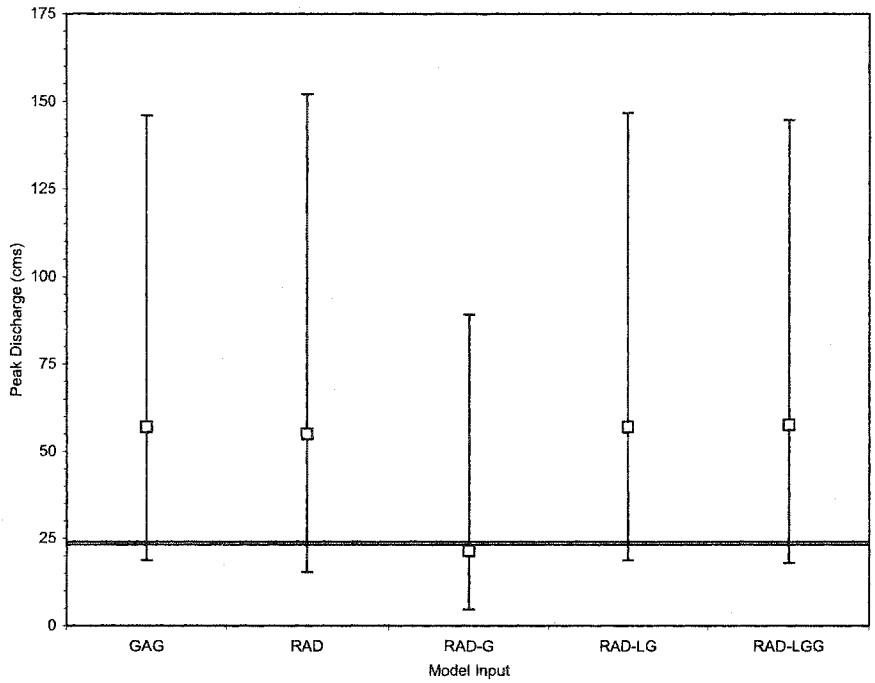


FIG. 4.23. As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case.

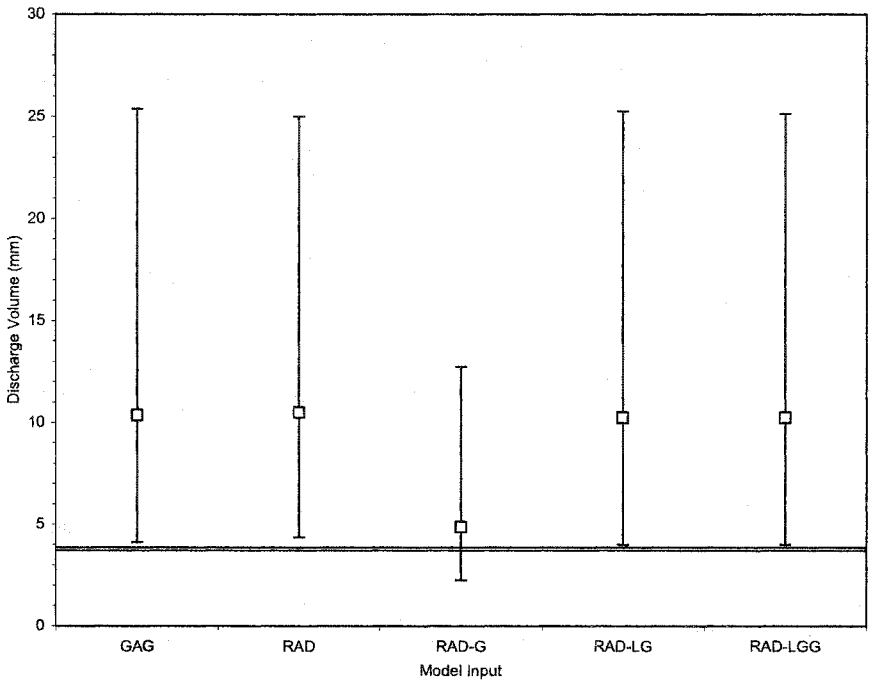


FIG. 4.24. As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 03 December 2002 case.

ensembles produced from GAG, RAD, RAD-LG, and RAD-LGG inputs yield very similar simulation bounds and medians for Peak. While the 90% simulation bounds include the observed Peak for these inputs, there appears to be a high bias with the model inputs. The RAD-G input, on the other hand, results in less uncertainty and the median Peak prediction aligns very close to the observed value.

Simulation bounds for the predicted Volume using the RAD-G input (Fig. 4.24) include the observed Volume. This is not the case with the ensembles that use inputs from the other rainfall algorithms. Their uncertainty bounds are relatively wide, but biased so high that the 90% simulation quantiles do not include the observed Volume. These results point to the RAD-G algorithm as providing the most accurate inputs to the *Vflo* ™ model when the entire parameter space is considered. Artifacts noted in the RAD algorithm (Fig. 4.17b) reveal possible calibration errors with the KTLX radar resulting in erroneously high accumulations. In spite of this apparent bias, gauge adjustments that rely only on gauges in close vicinity to the Blue River Basin do little to adjust the overestimated precipitation amounts. In fact, local gauge adjustments to the radar rainfall field yield products that are nearly indistinguishable from the original one. Evidently, there is good radar-gauge agreement at the handful of grid points surrounding the Blue River Basin, but these comparisons are not representative of the gridded rainfall over the basin. When over 100 radar-gauge comparisons are

made over the entire state, they indicate the radar-based accumulations are too high. The RAD-G product is thus biased low compared to all the other estimators. In this case, it turns out to be the most accurate input for hydrologic predictions of Peak and Volume variables. It is noted in this case how the Time predictions are least accurate using the ensemble that produces the most accurate Peak and Volume predictions. This apparent paradox is examined further in section 4.1.4 below.

4.1.4. *Summary of Hydrologic Evaluation*

The utilization of model parameter ensembles provides a comprehensive framework to evaluate the skill and uncertainty with hydrologic predictions that use different rainfall inputs. The model parameter ensembles include parametric uncertainty in the *Vflo* ™ model. Model inputs are evaluated from the hydrologic perspective in an objective and quantitative manner. Moreover, the parameter ensemble approach requires no prior modeling experience or calibration of model parameters in order to evaluate rainfall inputs. Time and Peak hydrologic variables are important for predicting the onset and magnitude of streamflow that may cause flash flooding. The time integrated volume of flow that exits the basin (Volume) is important for estimating the amount of water impacting downstream users of water such as irrigators, water delivery companies, and dam operators. This variable also sums the entire basin response throughout a storm event and is

thus a good indicator of the quality of rainfall inputs to the model. Biases in the initial inputs can be identified if other components of uncertainty in the modeling process (e.g., parametric) are taken into account. Three events are used to evaluate 9 different QPE algorithms that are input independently to the *Vflo* ™ model. Model parameter ensembles are created for each input in order to account for the uncertainty present in the parameters alone. No accounting of model structural errors or observational uncertainties is performed with this hydrologic evaluation. The observational uncertainty associated with streamflow measurements is smaller than the other error sources. As shown in section 4.2.2, the model structural uncertainty is also minimized for these three evaluation cases. The results from this study may be specific to the model and its application over the Blue River Basin. Moreover, only three case studies are used for the hydrologic evaluation. Nonetheless, important information is revealed about the model inputs, and the unique methodology proposed herein can be applied in a similar fashion for many more cases or for different environmental modeling systems altogether.

The statistical results from all three cases have been averaged and are summarized in Table 4.7. Ensemble mean *bias*es for predicted time of maximum discharge using all inputs are within 10% of observed Time values. The *RPS* skill value indicates the RAD algorithm produces the best time to peak predictions. The *bias*, *MAE*, *RMSE*, and *RPS* values all indicate the MS input results in the

TABLE 4.7. As in Table 4.1 but for average statistical evaluation of hydrologic predictions for all three cases combined.

TIME	Bias	MAE	RMSE	RPS
GAG	0.92	0.53	0.59	1.70
RAD	0.95	**0.41**	**0.51**	**0.83**
RAD-G	0.95	0.47	0.54	1.00
RAD-LG	0.95	0.46	0.55	1.04
RAD-LGG	0.94	0.44	0.55	0.93
MS	1.09	0.93	0.95	2.89
MS-G	**1.03**	0.60	0.62	1.96
MS-LG	0.92	0.54	0.56	2.07
MS-LGG	0.92	0.51	0.54	1.58

PEAK	Bias	MAE	RMSE	RPS
GAG	3.06	36.18	46.72	3.45
RAD	3.02	35.91	47.08	2.82
RAD-G	1.90	31.67	43.73	1.38
RAD-LG	3.10	38.11	50.40	3.05
RAD-LGG	3.52	58.98	75.58	3.98
MS	0.61	17.28	**19.18**	1.23

MS-G	**1.29**	**15.53**	22.06	**0.56**
MS-LG	2.15	28.30	38.66	2.37
MS-LGG	2.42	34.60	47.10	2.83

VOLUME	Bias	MAE	RMSE	RPS
GAG	3.38	7.02	8.82	4.31
RAD	3.40	7.73	9.62	3.91
RAD-G	2.56	7.10	8.73	3.39
RAD-LG	3.40	7.87	10.10	3.85
RAD-LGG	3.81	10.33	12.39	4.99
MS	**1.23**	**1.51**	**2.18**	**0.61**
MS-G	1.88	3.85	5.27	1.92
MS-LG	2.61	5.67	7.51	3.34
MS-LGG	3.02	7.10	8.84	4.28

worst Time predictions. The pdfs from all cases (Figs. 4.3, 4.11, and 4.19) show that there are two distinct modes of behavior for Time from all ensembles. The MS ensembles favor the modes that are out of phase with observed Time for both cases.

Parameter settings in both modes are examined to determine the nature of the bimodal behavior. As expected, the early mode is associated with smaller scalars applied to Manning roughness coefficients (n). Smoother surfaces result which leads to early timing of peak discharge. Bimodal behavior in the predicted pdfs may result from a nonlinear response in Time predictions to a given parameter. The scalars chosen to multiply the n parameter are distributed linearly. The scalars applied to the saturated hydraulic conductivity parameter (K) show no sensitivity to either mode. Scalars applied to the initial soil saturation parameter (θ) are quite different for the two different modes. Smaller relative minima in the predicted Time pdfs all correspond to members that use θ values of 100%. Saturated soils, however, do not necessarily lead to an expected early or late timing of the peak discharge from the model. Fig. 4.3, for example, shows Time predictions using the RAD-G inputs with 100% saturation favor the early mode, while predictions from all other QPE algorithms tend toward the later mode. On the other hand, when the RAD-G product is used as input for the 03 December 2002 case, the secondary maximum in the pdf (corresponding to 100% initial soil saturation; Fig. 4.19) occurs at a later time period. Precipitation inputs with

different temporal and spatial characteristics interact with complex parameter settings to cause members with 100% soil saturation to deviate nonlinearly from the other members that use unsaturated initial soil conditions. This deviation is not consistent, however, and may result in either later or earlier Time predictions. In any case, the bimodal shapes of the Time pdfs are explained by nonlinear sensitivity of Time predictions to initial soil saturation values of 100%.

Table 4.7 shows the ensemble that produces the most skillful Peak predictions results from inputs from the MS-G rainfall algorithm. Evidently, mean field bias adjustments applied to the MS algorithm result in more skillful hydrologic predictions of peak discharge. The RAD product also benefits from mean field bias adjustments. The *RPS* for the RAD product improves from 2.82 to 1.38 after mean field bias adjustments have been applied. Gauge adjustment strategies that rely more heavily on individual gauge amounts (i.e., -LG) result in degraded hydrologic predictions. In fact, this evaluation indicates that Peak predictions using RAD and MS with no gauge adjustment are more skillful than those that rely on local bias adjustment schemes (RAD-LG, RAD-LGG, MS-LG, MS-LGG) or rain gauge data alone (GAG). This hydrologic evaluation indicates the model-preferred gauge adjustment strategy is a mean field bias adjustment for the most likely Peak predictions on the Blue River Basin. This gauge adjustment strategy maintains spatial details in the precipitation fields while calibrating them using dozens of rain gauges in a domain-wide sense.

Bimodal behavior is also noted with Peak predictions for all ensemble pdfs (see Figs. 4.4, 4.12, and 4.20). Parameter settings associated with members belonging to the lower density, higher Peak mode were found to have initial soil saturation settings of 100%. Once again, the bimodal behavior of the pdfs is explained by a nonlinear response in Peak predictions to 100% soil saturation. As opposed to Time predictions, the result of this parameter setting is consistent and thus can be expected. All Peak predictions are much higher as compared to the rest of the members when the model soils are initially saturated. The hydrologic analysis offers clues about potential model structural uncertainties, especially in regards to the infiltration physics used in the *Vflo* ™ model. Identification of additional model uncertainties is explored in depth in section 4.2.2.

Table 4.7 indicates a significantly improved *RPS* for predicted Volume is associated with model parameter ensembles that use MS rainfall inputs. Predictions of Volume are thus most skillful with this input when model parametric uncertainty is included. This brief hydrologic analysis indicates that the most accurate QPE from the modeling perspective is the MS algorithm *with no gauge adjustment*. These initial results differ from QPE algorithms that are traditionally developed and compared against independent rain gauge amounts. In this case, the MS-LG and RAD-LG algorithms are forced to match individual gauge observations. However, they do not result in improvements at the larger hydrologic scale of a basin. The analysis shows that Mesonet gauges, while

considered to be a dense network (115 stations in the state of Oklahoma), often do not capture the details in the rainfall field needed for accurate hydrologic predictions. A slight degradation in hydrologic predictions of Volume occurs when gauge adjustments are applied to the MS algorithm. However, the RAD product benefits from a mean field bias adjustment. Note that the RAD-G product maintains the spatial characteristics of the RAD algorithm, but biases them depending on mean, statewide radar-gauge comparisons. The hydrologic evaluation has indicated that details in the spatial patterns of precipitation sensed by radar and satellite are indeed important for hydrologic purposes. Degraded Volume predictions result when these details are smoothed out by interpolated rain gauge observations.

Similar to Peak predictions, Volume predictions from all ensembles have a bimodal shape to their pdfs. Initial soil saturation values of 100% are coincident with lower density, higher Volume predictions. Hydrologic predictions from the *Vflo* ™ model are quite sensitive to this particular parameter setting. Volume predictions increase nonlinearly resulting in the bimodal pdf behavior. It is assumed that model structural uncertainties are minimal in this hydrologic analysis of model inputs while parametric uncertainty has been accounted for. As it turns out, all ensembles are affected similarly by the nonlinearity in Volume predictions to 100% soil saturation settings. Section 4.2.2 addresses the identification of model structural errors in detail.

4.2. Evaluation of the Characteristics of Uncertainty

This section addresses the second science question posed in this study: What are the predictability and limits-to-prediction in the hydrologic system? Chapter III refines this question further by proposing methodologies to illuminate several *characteristics* of uncertainty. The first task evaluates the sensitivity of hydrologic predictions to perturbed rainfall inputs. Are hydrologic predictions sensitive to a doubling of rainfall inputs? How do uncertainty bounds associated with the predictions respond to perturbed rainfall inputs? Results from this sensitivity study also apply to the third science question posed in this study: How do rainfall errors propagate to errors in streamflow predictions? The sensitivity study highlights the propagation characteristics of model input errors on streamflow predictions.

The hydrologic evaluation in section 4.1 reveals some peculiarities in predicted pdfs of model parameter ensembles. Two modes of behavior are shown to affect all ensembles that use different rainfall inputs. The presence of model parametric or structural errors is suggested but not yet shown explicitly. Another characteristic of uncertainty, the identification of model structural errors, is revealed in this study. Ensembles are constructed to include the combined effect of uncertainties in model inputs and model parameters. Observational uncertainties can be neglected, which leaves any residual error a result of uncertainties in the model structure or inadequate physical parameterizations.

The physical process for this uncertainty is highlighted and supported with evidence from independent measurements.

No known studies have attempted to estimate the *total* prediction uncertainty associated with a hydrologic forecast. Several techniques have been devised to primarily account for uncertainty in the model parameters. In section 4.1, for example, ensembles are created from the physical ranges associated with each parameter. The derived pdfs, however, are conditioned on perfect model inputs, physics, and observations. The first condition is rarely satisfied because different QPE algorithms are used as inputs and are found to have varying degrees of accuracy. The total prediction uncertainty is estimated for a case where observational and model structural uncertainties are minimized. Combined ensembles are created to include uncertainty in the inputs and parameters. This procedure provides for accurate estimation of predicted pdfs. This allows for pdfs to be utilized in decision-making that relies on specific probabilities. This ensemble prediction system will be useful for real-time applications if it is possible to perform the number of simulations needed for accurate estimation of the pdf. Many of the combined ensembles require thousands of simulations. The final study evaluates the minimum number of ensemble members needed to achieve the skill scores obtained with the full-blown ensemble. This information will provide for an estimate, on average, of the minimum number of members needed to optimize computational resources.

4.2.1. *Propagation of Uncertainty from Inputs to Hydrologic Predictions*

The sensitivity of hydrologic forecasts to perturbed rainfall inputs is examined using the 12 November 1994 case (Table 3.1). For this case alone, a version of the *Vflo* ™ model is used that assumes the channel side slopes are constant. The hydrograph for this case is shown in Fig. 4.25. Rainfall estimates are provided by the Arkansas Basin River Forecast Center's P1 algorithm. Model parameter ensembles are created the same as in the hydrologic evaluation (see section 4.1). In this case, model inputs have been scaled from their deterministic values, and thus aren't independent from each other as they are in section 4.1. The following scalars are used to produce five perturbations of model inputs: 0.5, 0.75, 1.0, 1.33, and 2.0 (corresponding to RAIN*0.5, RAIN*0.75, RAIN*1.0, RAIN*1.33, and RAIN*2.0 henceforth). Predicted ensembles are then compared to observed hydrologic variables following the analysis employed in section 4.1.

Table 4.8 shows a statistical comparison between predictions and observations of Time, Peak, and Volume. Table 4.9 indicates all *RPS* values being compared are different at the 99% statistical significance level. The time of peak discharge is predicted most skillfully using rainfall inputs that have been halved. *Bias*es of the ensemble means indicate the Time is predicted too early from the other perturbed inputs. Predictions of Peak have the lowest *RPS* value when inputs from RAIN*0.75 are used. Intuitively, the RAIN*.05 ensemble has a

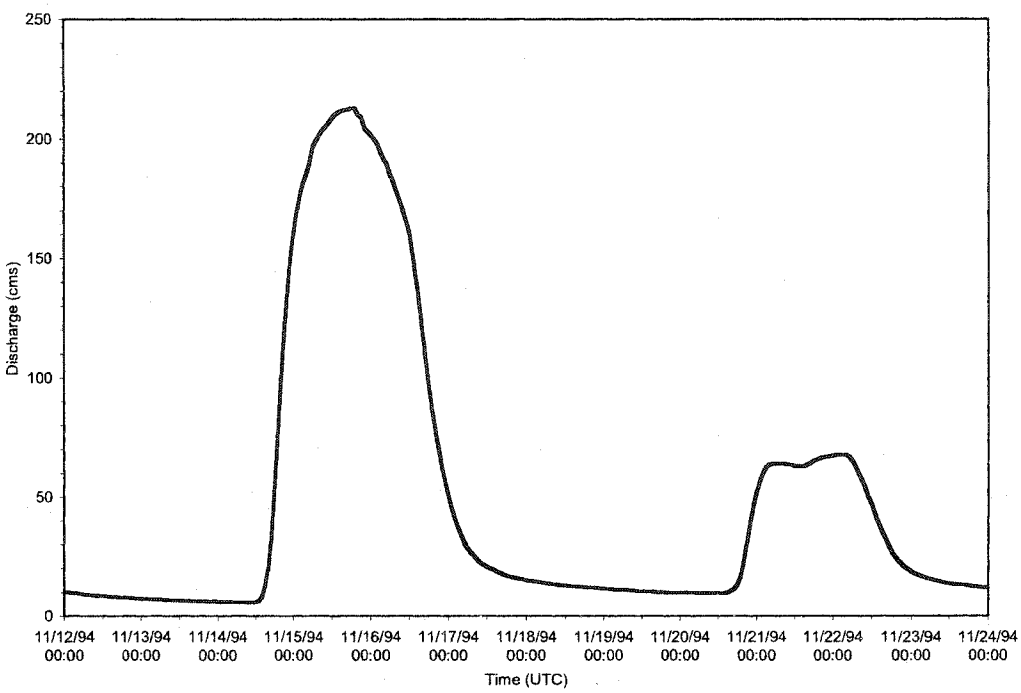


FIG. 4.25. As in Fig. 4.2 but for observed hydrograph for 12 November 1994 case.

TABLE 4.8. Statistical evaluation of hydrologic predictions for the 12 November 1994 case. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). The rainfall inputs are perturbed by scalars as noted in first column. See section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.

TIME	Bias	MAE	RMSE	RPS
RAIN*0.5	**1.01**	**0.35**	0.43	**0.55**
RAIN*0.75	0.95	0.37	**0.42**	1.05
RAIN*1.0	0.92	0.43	0.50	1.47
RAIN*1.33	0.89	0.55	0.59	2.14
RAIN*2.0	0.87	0.67	0.70	3.03

PEAK	Bias	MAE	RMSE	RPS
RAIN*0.5	0.42	124.52	133.90	1.79
RAIN*0.75	**0.95**	**85.34**	**101.05**	**0.51**
RAIN*1.0	1.60	155.55	205.59	1.15
RAIN*1.33	2.60	345.16	426.55	3.19
RAIN*2.0	4.97	855.03	970.12	5.75

VOLUME	Bias	MAE	RMSE	RPS
RAIN*0.5	0.35	26.92	27.18	2.41
RAIN*0.75	0.67	13.47	14.70	1.00
RAIN*1.0	**1.04**	**6.66**	**8.10**	**0.16**
RAIN*1.33	1.58	23.97	26.15	1.74
RAIN*2.0	2.77	72.83	74.45	5.63

TABLE 4.9. As in Table 4.2 but for significance levels of RPS differences for the 12 November 1994 case.

TIME	RAIN*0.5	RAIN*0.75	RAIN*1.0	RAIN*1.33	RAIN*2.0
RAIN*0.5	0.12	0.99	0.99	0.99	0.99
RAIN*0.75		0.07	0.99	0.99	0.99
RAIN*1.0			0.05	0.99	0.99
RAIN*1.33				0.06	0.99
RAIN*2.0					0.06

PEAK	RAIN*0.5	RAIN*0.75	RAIN*1.0	RAIN*1.33	RAIN*2.0
RAIN*0.5	0.06	0.99	0.99	0.99	0.99
RAIN*0.75		0.14	0.99	0.99	0.99
RAIN*1.0			0.06	0.99	0.99

RAIN*1.33				0.03	0.99
RAIN*2.0					0.06

VOLUME	RAIN*0.5	RAIN*0.75	RAIN*1.0	RAIN*1.33	RAIN*2.0
RAIN*0.5	0.14	0.99	0.99	0.99	0.99
RAIN*0.75		0.10	0.99	0.99	0.99
RAIN*1.0			0.23	0.99	0.99
RAIN*1.33				0.06	0.99
RAIN*2.0					0.09

low *bias* when considering the mean, while ensembles that use inputs from the larger scalars (i.e., RAIN*1.0, RAIN*1.33, RAIN*2.0) overforecast the peak discharge. The ensemble using unperturbed rainfall inputs (i.e., RAIN*1.0) produces the most skillful Volume predictions. There isn't a single perturbation that satisfies all three hydrologic objectives: Time, Peak, and Volume. Lower rainfall amounts result in the best Time predictions, while rainfall scalars of 0.75 and 1.0 produce the best Peak and Volume predictions respectively. Explanations for this discrepancy lie in the modified channel hydraulics that are used in this sensitivity study alone.

Gaussian kernel density estimation is used to produce pdfs for all ensembles. These pdfs reveal subtleties, such as bimodal pdf shapes, that aren't

as obvious in the statistical analysis. The pdfs of hydrologic predictions from each ensemble are shown in Figs. 4.26-4.28. The Time pdfs reveal a bimodal shape to each curve (Fig. 4.26). The secondary maximum was discovered in section 4.1 to be a result of model sensitivities to 100% soil saturation. Time

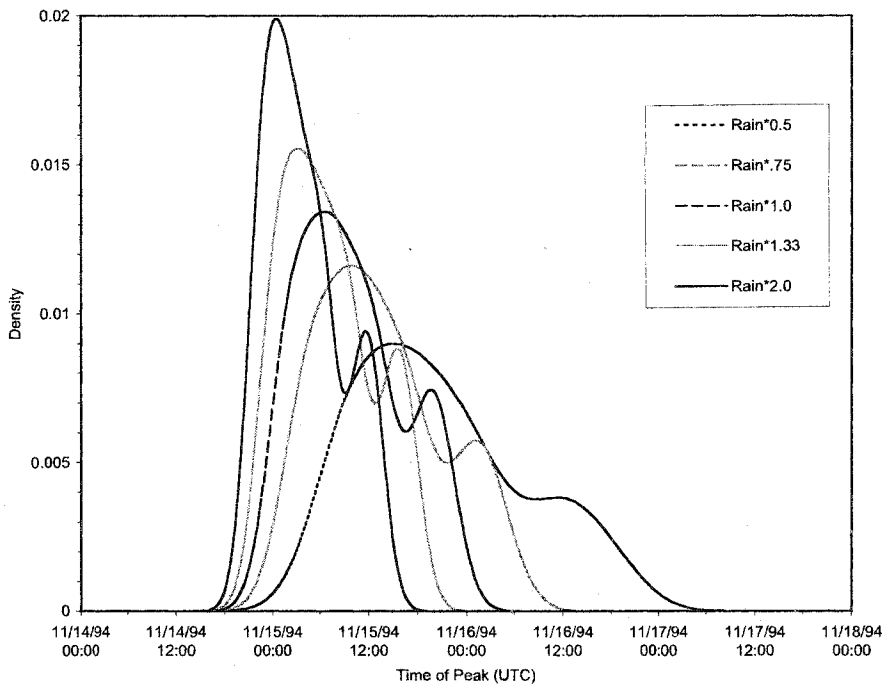


FIG. 4.26. As in Fig. 4.3 but for pdfs of the predicted time of maximum discharge (Time) for the 12 November 1994 case. Observed Time is 2000 UTC 15 November 1994.

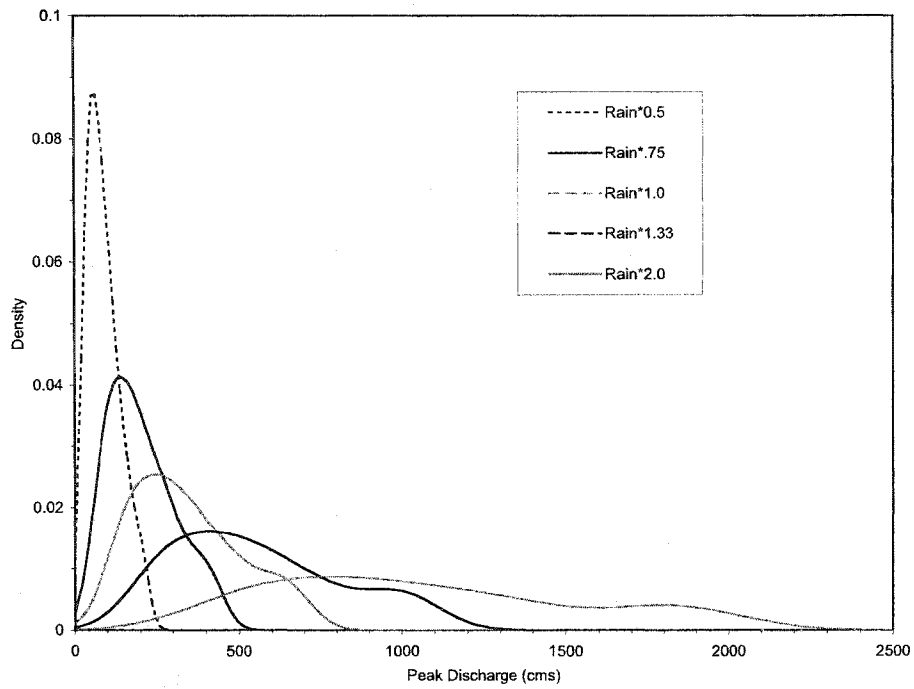


FIG. 4.27. As in Fig. 4.3 but for pdfs of the predicted magnitude of maximum discharge (Peak) for the 12 November 1994 case. Observed Peak is 215.2 cubic meters per second (cms).

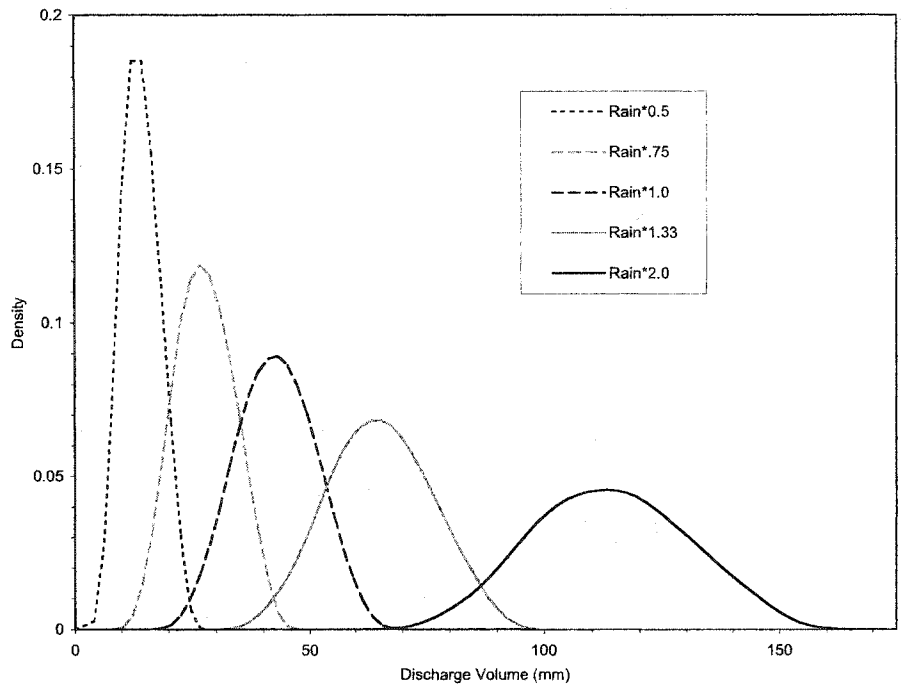


FIG. 4.28. As in Fig. 4.3 but for pdfs of the predicted time-integrated discharge volume normalized by the basin area for the 12 November 1994 case. Observed Volume is 41.1 mm.

hydrologic objectives. In this case, the rainfall "algorithms" are dependent on one another, so the predicted medians are plotted as a function of rainfall multiplier and connected by lines. Figure 4.29 shows the 90% simulation bounds encompass the observed Time only when rainfall scalars of 0.5 and 0.75 are used. predictions can occur either earlier or later than the rest of the members that use unsaturated initial soil conditions. In this case, the secondary peak occurs at later times, meaning 100% soil saturation results in a delayed time to peak for all ensembles. Lighter rainfall amounts (e.g., RAIN*0.5) also result in later, more correct Time predictions. The distributions of Peak predictions (Fig. 4.27) reveal the anticipated behavior of lower rainfall inputs (e.g., RAIN*0.5) resulting in smaller Peak predictions, while larger inputs (e.g., RAIN*2.0) produce much higher Peaks. There is a slight indication of bimodal behavior in the pdfs. The secondary bumps are all coincident with relatively higher Peaks, indicating members with 100% soil saturation produce anomalously high Peak predictions. Volume predictions from the ensembles show a similar dependence on rainfall scalar as with Peak predictions (Fig. 4.28). Higher rainfall scalars result in much higher predictions of Volume. In this case, however, the bimodal shape to the pdfs is not apparent with the Volume predictions.

Propagation characteristics of hydrologic predictions to rainfall perturbations are shown with 90% simulation bounds derived from the pdfs. Figures 4.29-4.31 show ensemble spreads and measures of central tendency for all

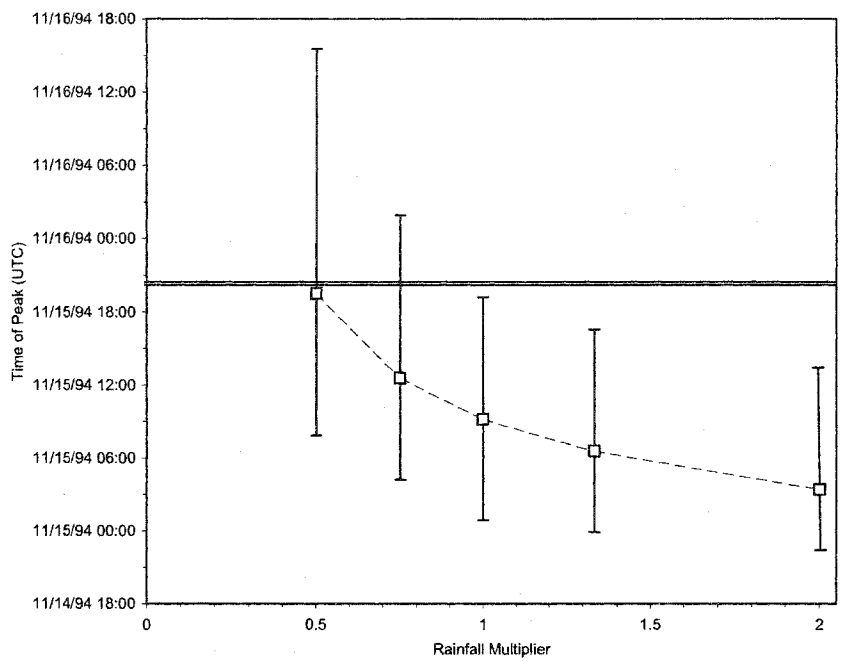


FIG. 4.29. As in Fig. 4.6 but for simulation bounds of the predicted time of maximum discharge (Time) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.

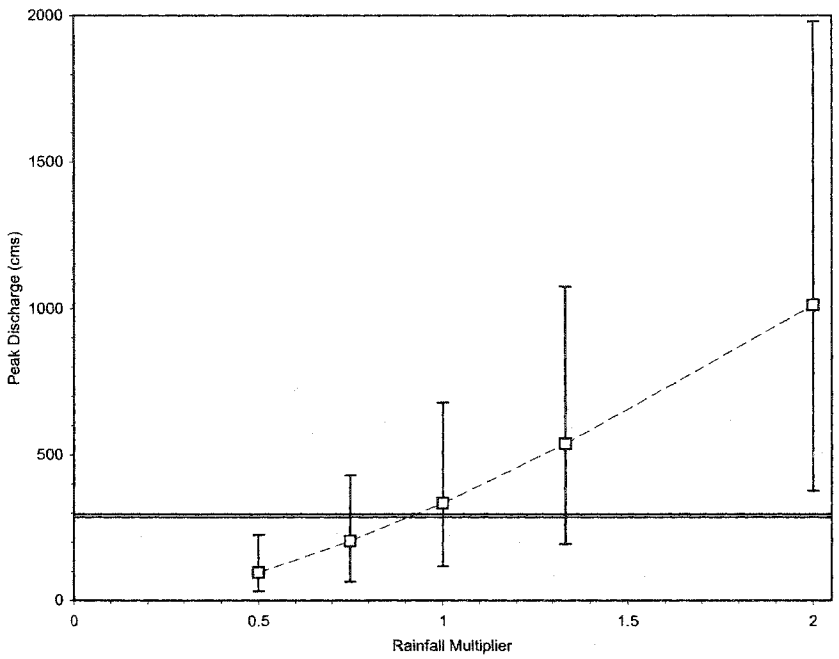


FIG. 4.30. As in Fig. 4.6 but for simulation bounds of the predicted magnitude of maximum discharge (Peak) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.

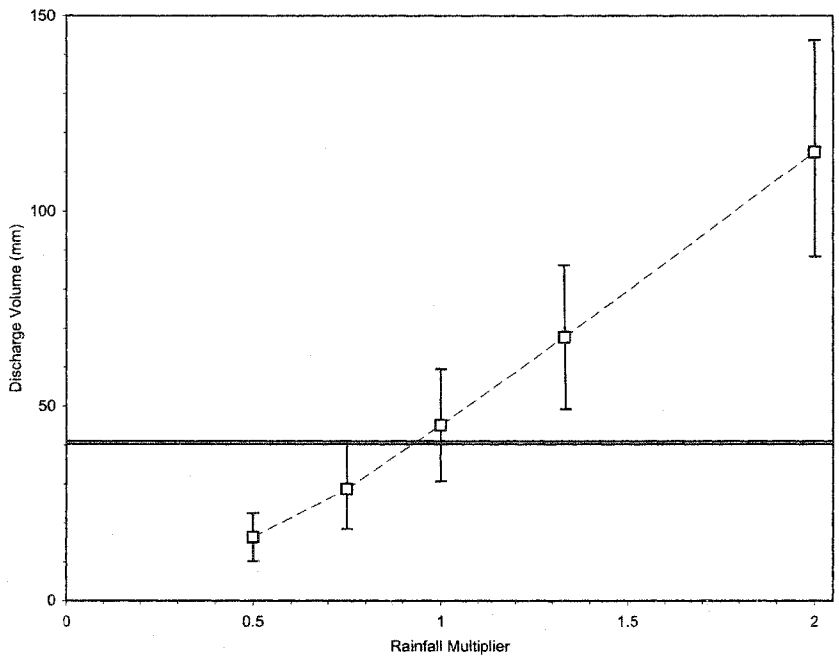


FIG. 4.31. As in Fig. 4.6 but for simulation bounds of the predicted time-integrated discharge volume normalized by the basin area (Volume) for the 12 November 1994 case. Model inputs are perturbations of the given rainfall estimates. Scalars used to perturb the inputs are indicated on the x-axis.

As rainfall magnitudes increase, peak discharges exit the Blue River Basin at earlier times. Higher flow velocities result from larger hydraulic radii (resulting from higher flow depths) from Manning's equation (2.7). Also, the shape of the curve connecting the ensemble medians should asymptotically approach the case start time if larger rainfall scalars were employed.

The unexpected result from Fig. 4.29 is the apparent "preference" the modified version of the *Vflo* ™ model has for lighter rainfall amounts in terms of timing. The amount of discharge flowing through the basin for this event is nearly an order of magnitude larger than the other case studies (Table 3.1). In a modified formulation of the model, the side slopes of the channels are defined to be constant at a given cross-section for all flow depths. In reality, the side slopes become much flatter at high flow depths, basically representing the flood plain adjacent to the riverbed. Thus, for extreme events water leaves the main channel and flows over a vegetated, rough flood plain. The observed Peaks are observed to arrive later at the basin outlet as a result of this. The analysis of error propagation through the model has highlighted an error in regard to specification of channel side slopes. This structural or parametric error has been mitigated using rating curves for all other uses of the *Vflo* ™ model in this study.

Predictions of peak discharge appear to be quite sensitive to perturbations in rainfall inputs (Fig. 4.30). Intuitively, Peak predictions increase with larger rainfall scalars. In addition, the spread of the ensembles increases dramatically

with higher rainfall amounts. Simulation limits for ensembles using rainfall scalars of 0.75, 1.0, and 1.33 encompass the observed Peak value, and are thus deemed behavioral. Unperturbed rainfall inputs (i.e., RAIN*1.0) result in an ensemble of predictions that include the observed Peak. It is indicated in Table 4.8, however, that the best *RPS* skill value is assigned to the ensemble using RAIN*0.75 inputs. Once again, it appears that the channel-modified version of the *Vflo* ™ model has a slight preference for lighter rainfall when predicting Peak discharge. Channel misspecifications are believed to result in erroneously early Time predictions *and* high Peak predictions. With this extreme event, water leaves the main channel and flows over the adjacent flood plain. This surface is much rougher than in the channel and results in hydrographs that are less peaked, and the peaks are delayed. The channel cross sections are not properly represented as a function of height in this version of the *Vflo* ™ model, which results in erroneous Time and Peak predictions. These model structural errors, however, should have little impact on Volume predictions. Improved channel hydraulics in the *Vflo* ™ model utilize rating curves to simulate the channel geometry as opposed to modeling it explicitly. Rating curves are used for all other experiments in this study. Thus, conclusions regarding channel misspecifications apply to this section alone.

Volume predictions are also sensitive to perturbations in rainfall inputs (Fig. 4.31). Higher Volumes result from increases in model inputs. The 90%

simulation bounds also widen as larger scalars are applied to the rainfall estimates. Volume predictions using ensembles with scaled rainfall of 0.75 and 1.0 include the observed Volume for this case. The median from the unperturbed ensemble lies closer to the observed which results in a better *RPS* skill value.

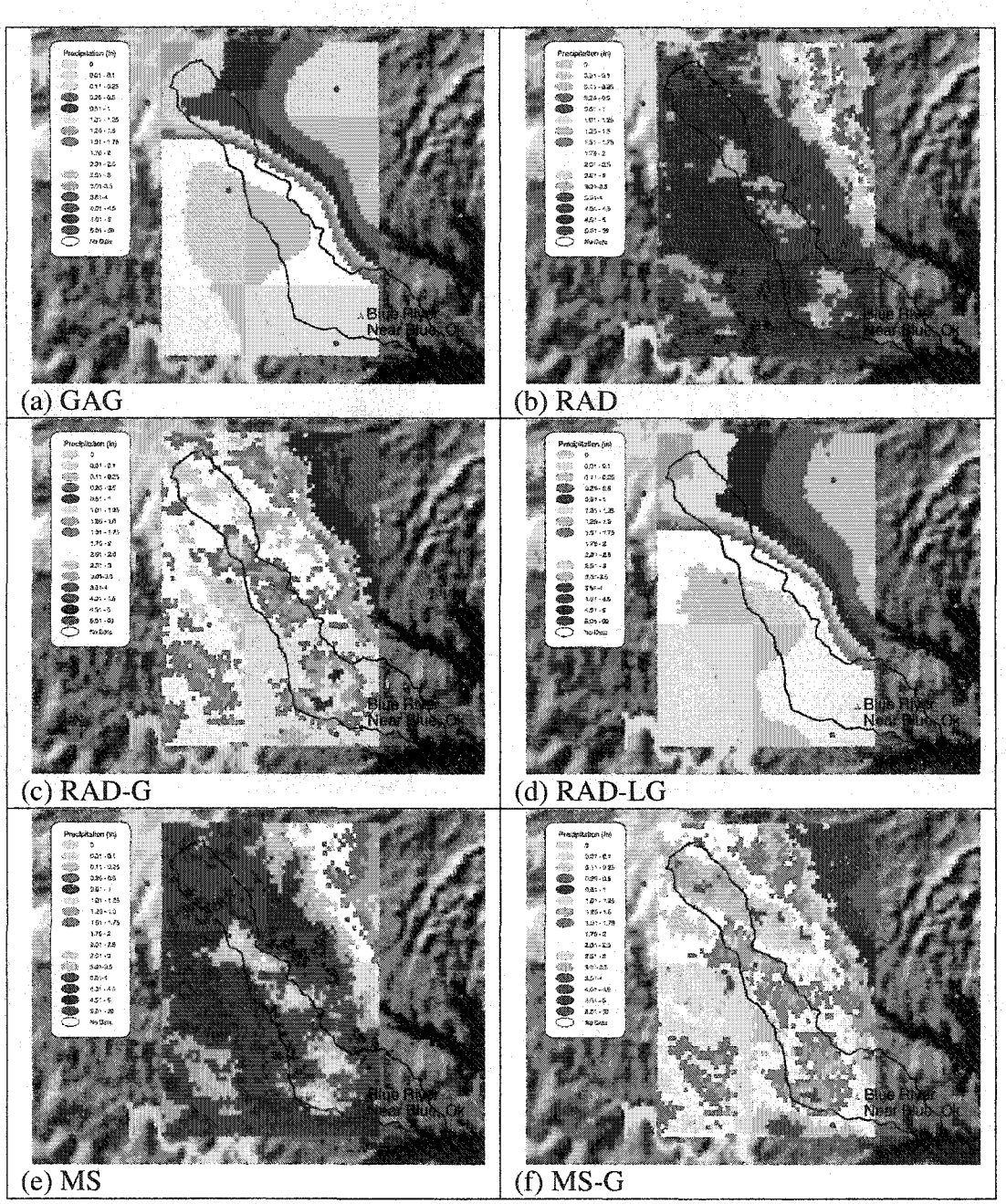
The third science question addressed in this study aims to understand how input errors propagate to hydrologic predictions. In the case of predicted time of maximum discharge, the relationship between input errors and hydrologic predictions is nonlinear, as indicated by the curve connecting the medians in Fig. 4.29. Large positive rainfall perturbations result in early Time predictions. However, the curves are nonlinear because the earliest possible Time prediction (associated with the largest scalars) is bound by the onset of precipitation. The latest possible Time prediction (associated with the smallest scalars) is theoretically bound by positive infinity. However, this situation will not occur because infinitessimal rainfall amounts will eventually result in no discharge at all. In any case, the curve connecting the simulation medians must asymptotically approach these bounds and thus take on a nonlinear shape.

Figures 4.30-4.31 show the relationships between rainfall perturbations on Peak and Volume predictions. The lines connecting the medians have a linear shape. It must be noted, however, that slopes of these lines are rather steep. When values on the y-axis in Fig. 4.30 are normalized by the observed peak discharge, a fitted line to the simulation medians has a slope of 2.87. This means

that a doubling of rainfall inputs results in Peak predictions that are approximately 5.74 times greater. For Volume, a fitted line to the medians has a slope of 1.62. This means that a doubling of rainfall inputs results in a Volume prediction that is approximately 3.24 times larger. The relationship between rainfall perturbations and hydrologic predictions of Peak and Volume are linear, but the slopes are steep indicating high sensitivity. This result is profound as it emphasizes the need to focus on the accuracy of QPE algorithms used in hydrologic modeling.

4.2.2. *Identification of Uncertainty in Model Structure*

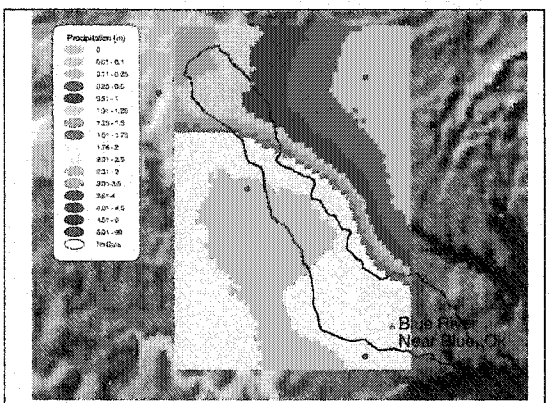
Structural uncertainties are identified when all other sources of uncertainty in the modeling process are accounted for, and simulations still do not give the observed behavior, i.e., it is the residual. Ensembles are constructed herein to account for uncertainties in the model parameters and inputs. Observational uncertainties can be safely neglected. Input uncertainties are handled by treating each QPE input as an equally likely estimator of the true rainfall field. Each QPE member uses rain gauges, radar, satellites, model data, or combinations in its estimation scheme. It is assumed that these members encompass the true spread of rainfall possibilities. Rainfall estimates for the 25 August 2002 case are shown in Fig. 4.32. Model parameters are perturbed within their physical bounds as is done in previous sections. Ensembles are constructed to include the combined uncertainty in inputs and parameters. A given ensemble is thus composed of the number of



(a) GAG (b) RAD

(c) RAD-G (d) RAD-LG

(e) MS (f) MS-G



(g) MS-LG

FIG. 4.32. As in Fig. 4.1 but for storm total precipitation plots for the 25 August 2002 case from the QPESUMS products: (a) gauge-only, (b) radar, (c) radar with mean field bias removed, (d) radar with local bias adjustment, (e) multisensor, (f) multisensor with mean field bias removed, and (g) multisensor with local bias adjustment.

parameter perturbations (125) multiplied by the number of available rainfall estimators (7). The observed hydrograph for the 25 August 2002 case is shown in Fig. 4.33.

Table 4.10 shows statistical results for combined (input + parameter) ensembles for four cases on the Blue River Basin. Peak and Volume predictions for the 25 August 2002 case are biased extremely high as noted in the *bias* statistic. Combined ensemble predictions for the October and December events do not show this overforecast problem. Moreover, the *MAE*, *RMSE*, and *RPS* values for Peak and Volume predictions with the 25 August 2002 case are much

TABLE 4.10. Statistical evaluation of hydrologic predictions from combined input-parameter ensembles for cases listed in the left column. The hydrologic variables being considered are the time of maximum discharge (Time), magnitude of maximum discharge (Peak), and time-integrated discharge volume normalized by the basin area (Volume). See section 3.2.2 for the statistical definitions. Numbers in boldface indicate the best agreement with observations.

TIME	Bias	MAE	RMSE	RPS
25-Aug-02	1.04	0.42	0.47	**0.77**
23-Oct-02	0.95	**0.42**	**0.47**	1.10
28-Oct-02	**1.01**	0.77	0.83	1.52
3-Dec-02	0.94	0.54	0.67	0.96

PEAK	Bias	MAE	RMSE	RPS
25-Aug-02	35.93	216.59	243.29	6.00
23-Oct-02	**1.39**	32.02	48.49	**0.64**
28-Oct-02	2.27	**24.05**	**33.16**	2.15
3-Dec-02	4.27	44.16	57.51	4.21

VOLUME	Bias	MAE	RMSE	RPS
25-Aug-02	24.75	38.00	43.72	6.00

23-Oct-02	**2.12**	6.99	9.69	**2.01**
28-Oct-02	2.66	**3.82**	**5.17**	3.11
3-Dec-02	4.31	8.27	10.29	5.02

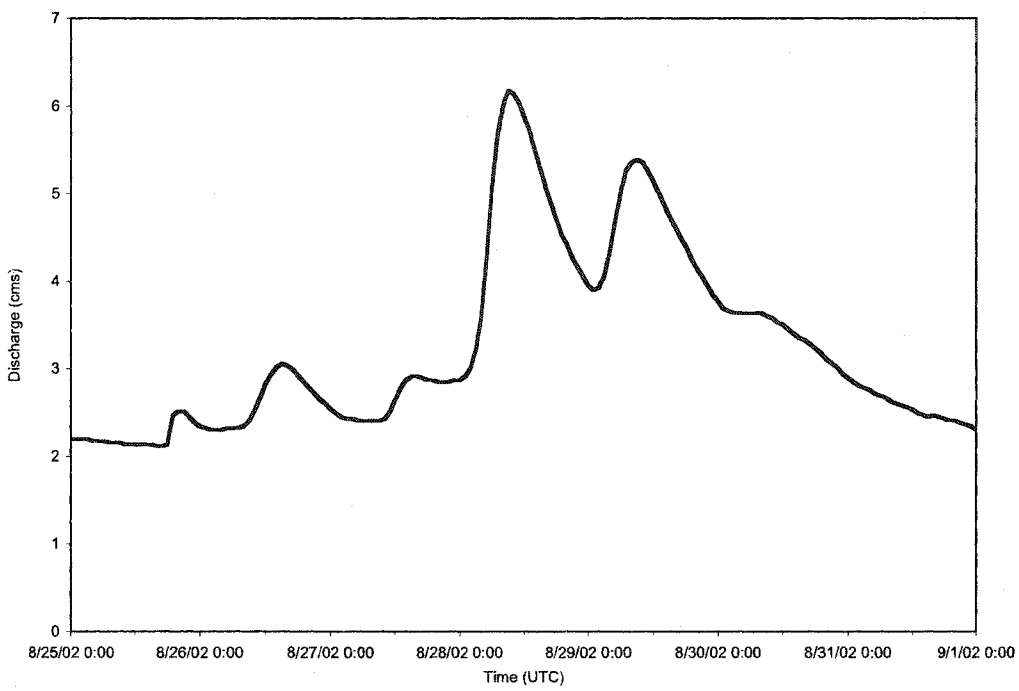


FIG. 4.33. As in Fig. 4.2 but for observed hydrograph for 25 August 2002 case.

worse than those for the other cases. The combined ensemble pdfs are shown in Figs. 4.34-4.36. Previously, the bumps in the curves were attributed entirely to model sensitivities to the soil saturation parameter. In this case, different modes may also be due to rainfall inputs. Multiple modes are possible when different regions of the phase space are preferred by different rainfall inputs and/or parameters. Figure 4.34 shows the pdf for Time predictions from the combined input-parameter ensemble has a bimodal shape. Predictions for Peak and Volume have trimodal distributions (Figs. 4.35-4.36) with smaller relative maxima occurring at larger Peak and Volume values.

The simulation bounds and medians for all cases listed in Table 4.10 are shown in Figs. 4.37-4.39. The combined ensembles provide accurate Time predictions for all cases. On the other hand, predictions of Peak and Volume by the combined input-parameter ensembles for the 25 August 2002 case are much greater than observed values. 90% simulation bounds for the other cases listed in Table 4.10 include the observed Peak and Volume amounts. However, the uncertainty limits for the 25 August 2002 case do not encompass observed Peaks and Volumes, and are thus deemed nonbehavioral. The combined ensembles include expected uncertainty in the inputs and parameters. The residual error in this summer case must be attributed to model structural errors or inadequate physical parameterizations. Additional data sets are introduced to identify the source of uncertainties in the *Vflo* ™ model.

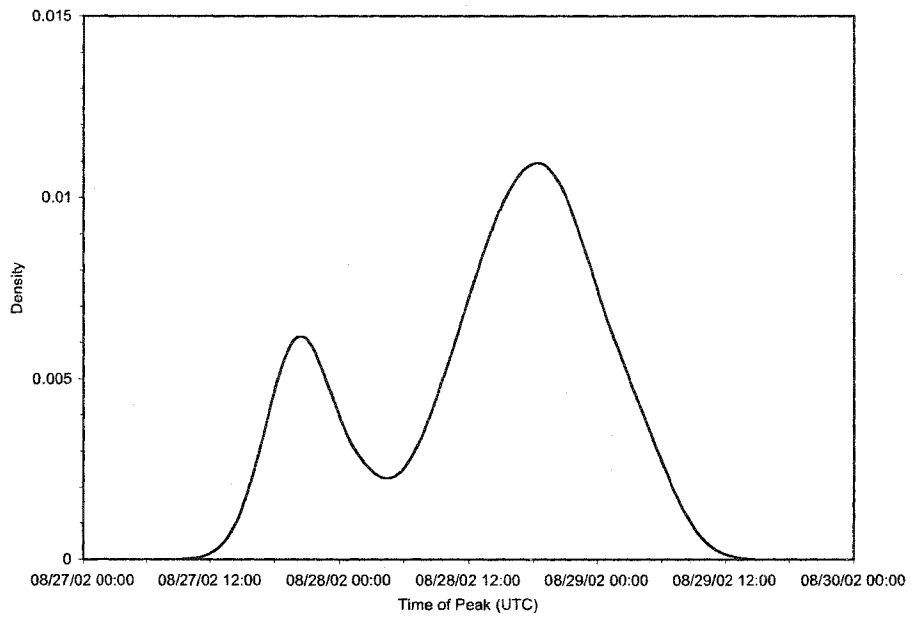


FIG. 4.34. Probability density function of the predicted time of maximum discharge (Time) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Time is 0900 UTC 28 August 2002.

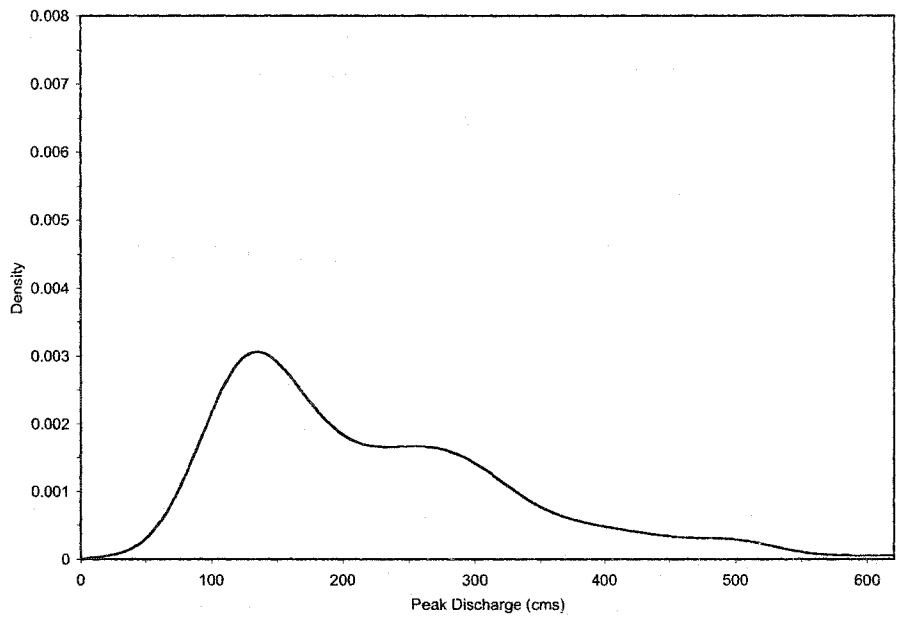


FIG. 4.35. Probability density function of the predicted magnitude of maximum discharge (Peak) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Peak is 6.2 cubic meters per second (cms).

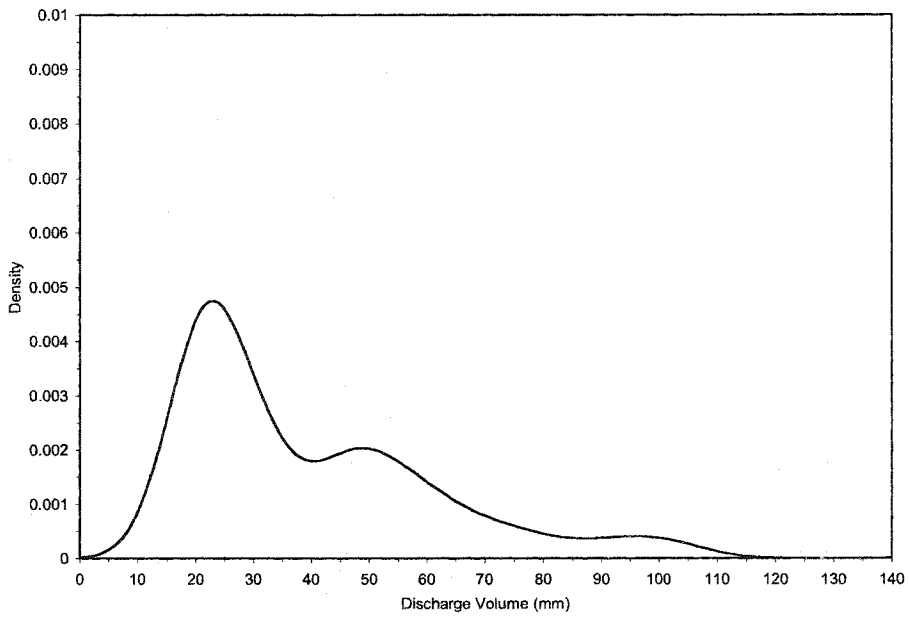


FIG. 4.36. Probability density function of the predicted time integrated discharge volume normalized by the basin area (Volume) for the 25 August 2002 case using a combined input-parameter ensemble. Observed Volume is 1.6 mm.

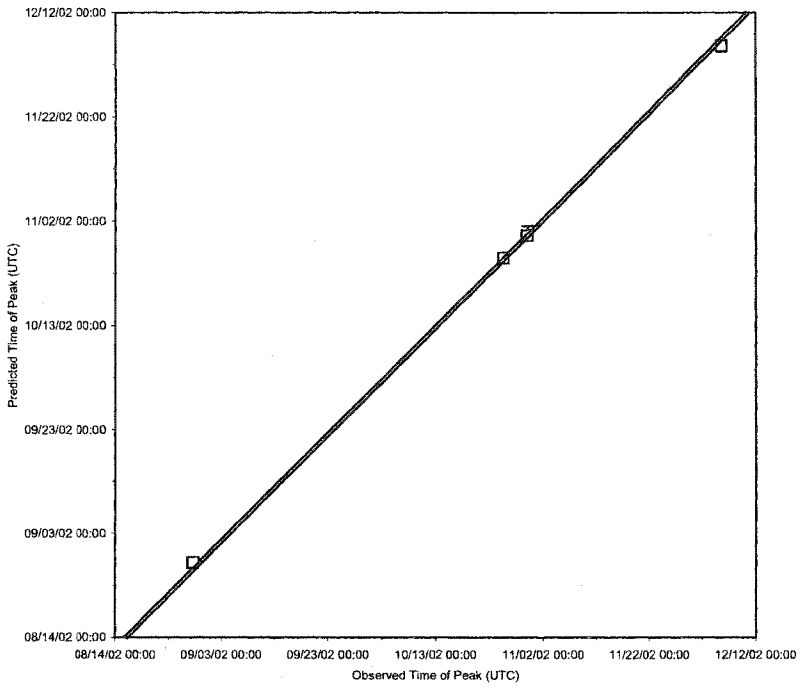


FIG. 4.37. Scatter plot of predicted time of maximum discharge (Time) versus observed Time for the cases listed in Table 4.10. The open boxes refer to the 50% quantile (median), while the bars correspond to the 5% and 95% quantiles. The diagonal double line is the 1:1 line.

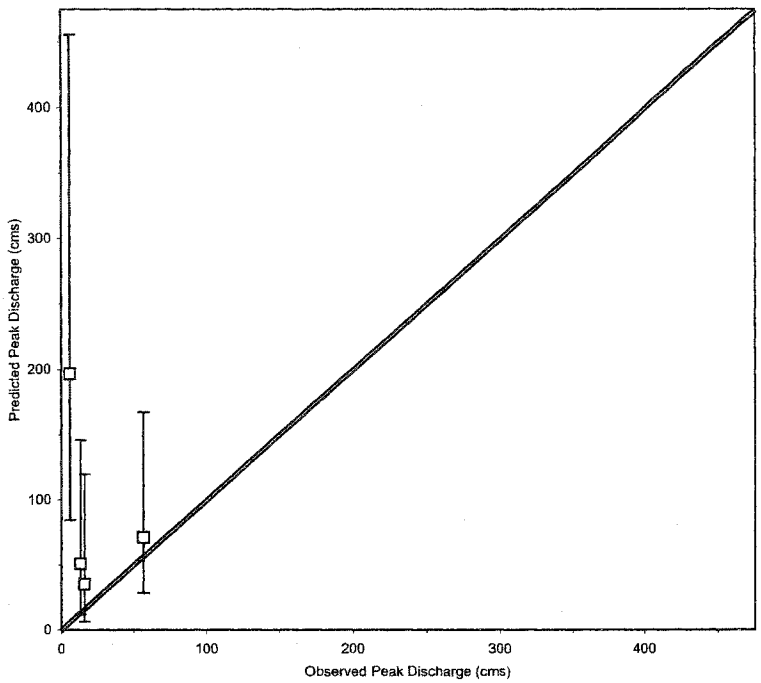


FIG. 4.38. As in Fig. 4.37 but for scatter plot of predicted magnitude of maximum discharge (Peak) versus observed Peak for the cases listed in Table 4.10.

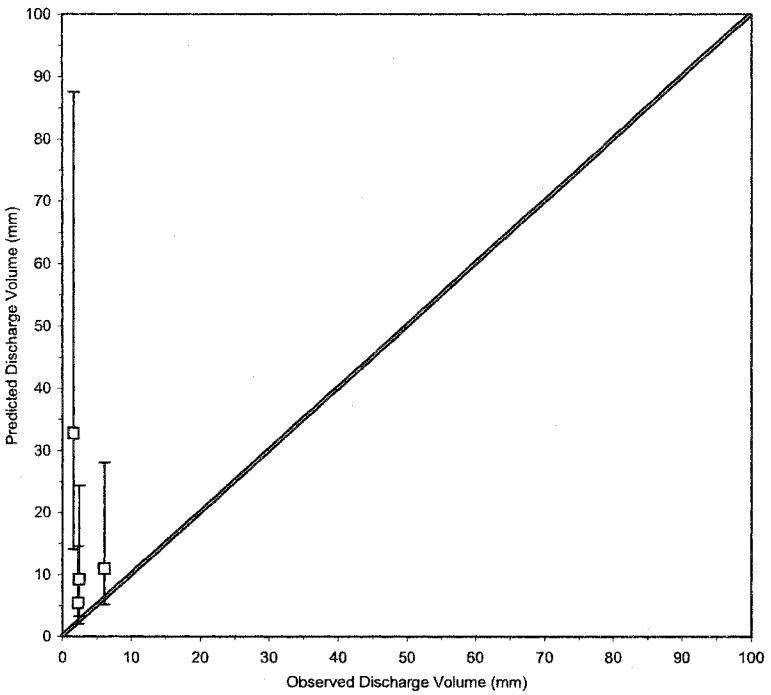


FIG. 4.39. As in Fig. 4.37 but for scatter plot of predicted time integrated discharge volume normalized by the basin area (Volume) versus observed Volume for the cases listed in Table 4.10.

The lack of significant runoff production (Fig. 4.33) for the 25 August 2002 case is not entirely intuitive given the relatively high rainfall amounts (Fig. 4.32) impacting the basin relative to the other cases (Figs. 4.1, 4.9, and 4.17). The main clue to this discrepancy lies in the time of year at which the events occur. An empirical study involving the relationship between rainfall and runoff on the Blue River Basin is undertaken to elucidate the apparent lack of runoff production during the summer months. Monthly streamflow observations and rain gauge observations taken at the Blue River Basin, near Blue, OK are collected from 1937 to 1997. Each monthly observation (61 total) is plotted in a scatter plot (not shown). A line is then fit to the climatological observations for each month. The slope of the line is a proxy for the rainfall-runoff relationship. Higher slopes mean more runoff results from a unit input of rainfall, while smaller slopes indicate little runoff is produced from the same unit input of rainfall. The slope of the fitted line is shown in Fig. 4.40. The trend of the slope indicates there is significantly different runoff production in the Blue River Basin from month-to-month. Independent observations of soil moisture conditions in the basin are introduced to identify the cause of the dependence of rainfall-runoff relationships on the month of the year.

The Oklahoma Mesonet has installed soil moisture sensors at several sites in the state of Oklahoma. Soil moisture observations have been collected on a daily basis at the Durant Mesonet site since 1997. This location resides within the

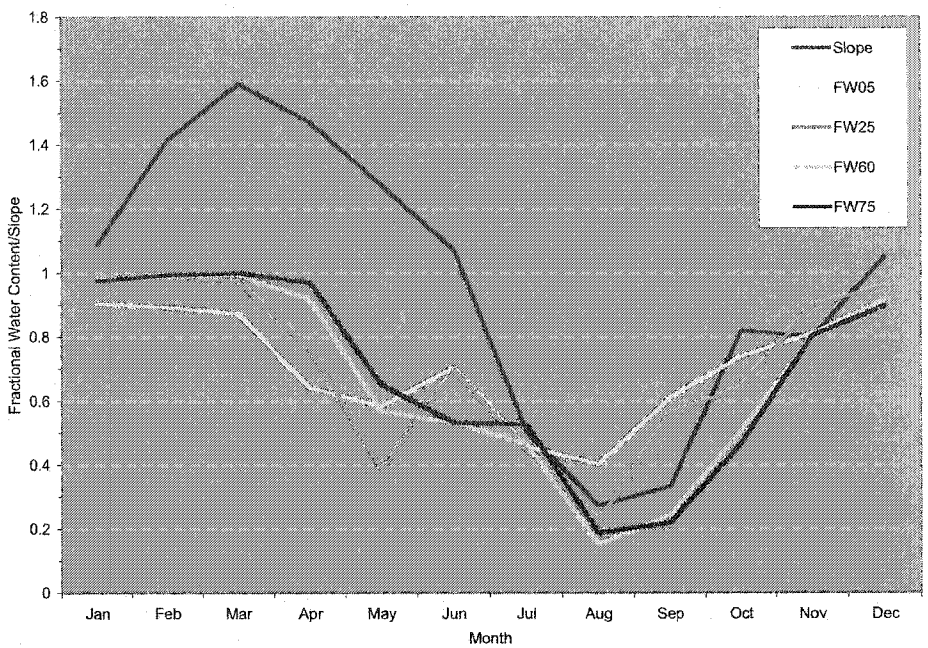


FIG. 4.40. Climatological trend of monthly rainfall-runoff relationships and soil moisture conditions on the Blue River Basin. The slope (blue curve) indicates the relationship between rainfall and runoff using 61 years of observations. The other curves (see legend) are monthly fractional water content values at depths of 5, 25, 60, and 75 cm. 7 years of soil moisture conditions from the Durant Mesonet site are used to construct the fractional water content climatologies.

Blue River Basin and is assumed to be representative of soil moisture conditions over the basin when several years of observations are considered. Daily measurements of fractional water content at 4 different depths (5, 25, 60, and 75 cm) are grouped into their respective months and averaged for the available, 7-year period. This short but representative soil moisture climatology is then plotted against the slope parameter that summarizes the rainfall-runoff relationship on the Blue River Basin (Fig. 4.40). The trends in the slope parameter and fractional water contents are well correlated. In fact, a correlation coefficient of 0.85 describes the linear relationship between the slope parameter and fractional water content at 75 cm. This indicates runoff production is very limited for this basin when deep-layer soils are dry. Recall, the initial soil saturation parameter is varied from 20% to 100% and thus represents dry conditions in the Green and Ampt model. Simulations still produce too much runoff which indicates the Green and Ampt model used in *Vflo*™ is not accounting for enough infiltration over dry, deep-layer soils.

Basara and Crawford (2002) discovered strong linear relationships between moderate to deep-layer soil moisture and atmospheric processes. Their relationships were found to be a result of the capability of vegetation to transport energy in the vertical, i.e., transpiration. This study indicates deep layer soil moisture is also correlated with surface hydrologic conditions, at least across monthly time scales. Perhaps the linearity between rainfall-runoff relationships

and soil moisture at deep layers is also influenced by moisture transport between the soil surface and the root zone.

The *Vflo* ™ model uses the Green and Ampt model (see Chapter II, section 2.2) for infiltration rates. Water becomes available for surface runoff in the event that rainfall rates exceed infiltration rates. Prior to this ponding time, all rainfall is infiltrated in the soil as an initial abstraction. Abstractions include the combined effects of infiltration as modeled by the Green and Ampt equation, depression storage on the surface, interception by vegetation, and enhanced infiltration caused by conduits for water flow in the soil surface. During the summer months, it is plausible that interception by vegetation and flow conduits caused by cracks result in higher infiltration rates than those predicted by the Green and Ampt equation. The combined ensemble methodology has identified a model structural error or inadequate parameterization of saturated hydraulic conductivity for this particular case. Future studies will examine larger parameter ranges for saturated hydraulic conductivity, include the wetting front suction parameter in the combined ensemble, and determine how initial abstractions can be handled directly as a boundary condition and included in the ensemble.

4.2.3. *Estimation of Total Prediction Uncertainty*

The total prediction uncertainty can be estimated if all sources of uncertainty in the modeling process are accounted for or are deemed negligible.

Section 4.2.2 demonstrates how combined ensembles are created to represent uncertainties in model inputs and parameters. Observational uncertainty of streamflow is much smaller in comparison and is negligible. Model structural uncertainty is more difficult to identify. It is treated as the residual source of uncertainty when all other possible roots are included in the ensemble. Model infiltration is discovered to be a significant error source, but is only problematic during low flow, summer months. This section demonstrates how the total prediction uncertainty is estimated for cases when the model structure is valid. Future studies should address how the Green and Ampt model can be modified or parameterized to adequately simulate infiltration during the summertime. Additional infiltration models may need to be considered.

Statistical results are obtained for the combined ensembles for the 23 October 2002, 28 October 2002, and 03 December 2002 cases (Table 4.10). The statistical significance of the *RPS* differences is not computed for these cases because the number of members comprising each ensemble varies. Pdfs of the predicted hydrologic variables are shown in Figs. 4.41-4.43 for the 23 October 2002 case. Figure 4.41 shows a flatter, more dispersive pdf for predicted Time. The extreme bimodal shape of model parameter pdfs noted in section 4.1 has been smoothed in the combined pdf. It should also be noted that the combined pdfs can be refined further by rejecting members that systematically don't produce observed system behavior. For example, members using initial soil saturation

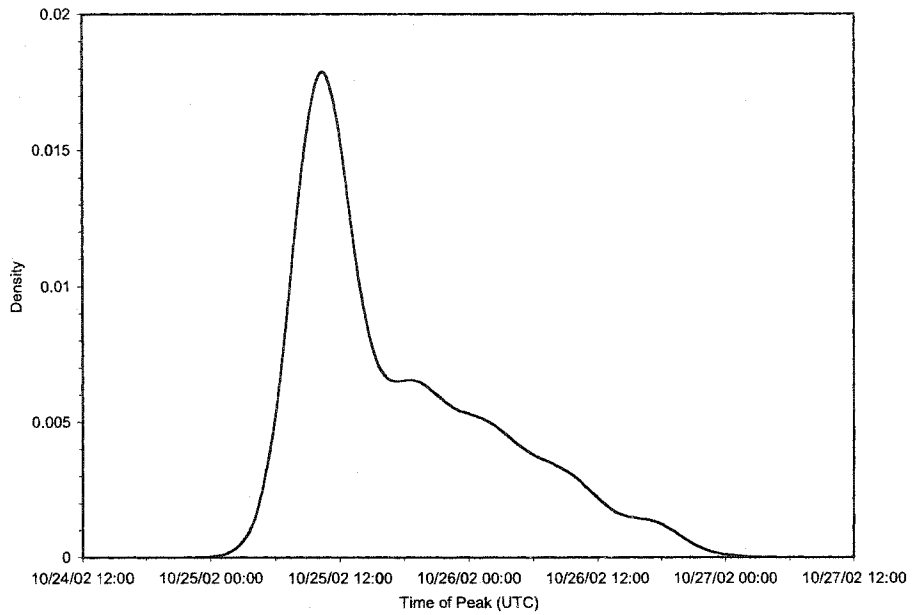


FIG. 4.41. Probability density function of the predicted time of maximum discharge (Time) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Time is 0000 UTC 26 October 2002.

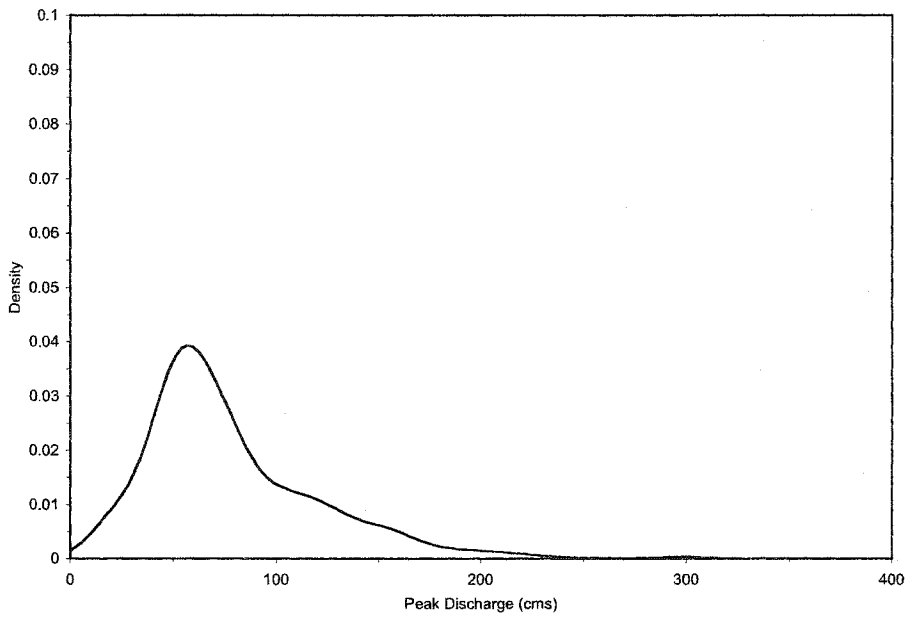


FIG. 4.42. Probability density function of the predicted magnitude of maximum discharge (Peak) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Peak 56.7 cubic meters per second (cms).

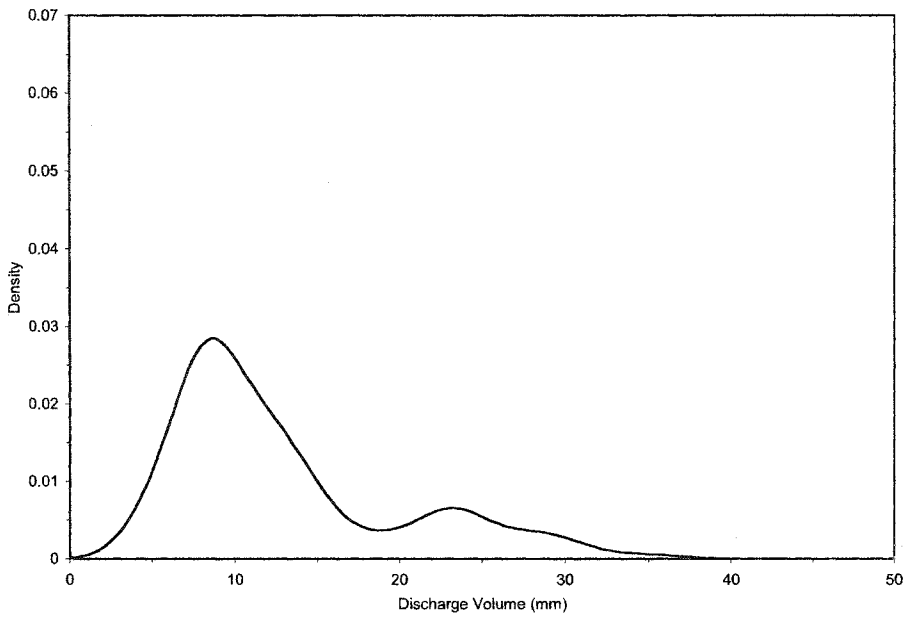


FIG. 4.43. Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 23 October 2002 case using a combined input-parameter ensemble. Observed Volume is 6.1 mm.

parameter settings of 100% produce unrealistic Time, Peak, and Volume predictions. These members can be rejected from the combined ensemble. Judicious selection of appropriate models will remove the bimodal shapes that are still apparent in the combined pdfs (Fig. 4.43). In addition, the hydrologic evaluation in Section 4.1 highlights specific QPE algorithms that are more likely to replicate the observed behavior. Future studies will weight these inputs appropriately in order to reduce the uncertainty bounds with the hydrologic predictions.

Figures 4.44-4.46 show the pdfs of the combined ensemble predictions for the 28 October 2002 case. The pdf for predicted Time (Fig. 4.44) is dispersive indicating several regions in the Time phase space are replicated by individual members comprising the combined ensemble. The predicted Peak and Volume pdfs (Figs. 4.45-4.46) are also more dispersive with a slighter indication of a false, secondary bump at higher predictions. Nonetheless, inclusion of QPE members in the combined ensemble produces predicted pdfs that better estimate the true pdf. Figures 4.47-4.49 show the pdfs of hydrologic predictions for the 03 December 2002. Similar observations are noted with these predicted pdfs. The predicted Time pdf (Fig. 4.47) in this case shows the input ensembles have strong preferences for either the early or late mode of behavior. Very few members predict Time values between the two peaks. This bimodal shape can be mitigated by excluding models employing the 100% soil saturation parameter setting. The

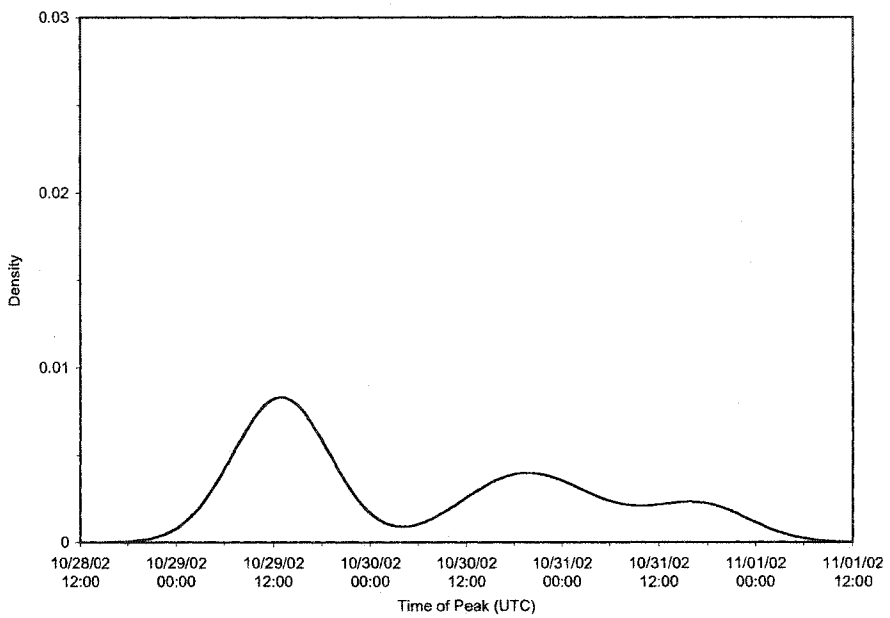


FIG. 4.44. Probability density function of the predicted time of maximum discharge (Time) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Time is 0600 UTC 30 October 2002.

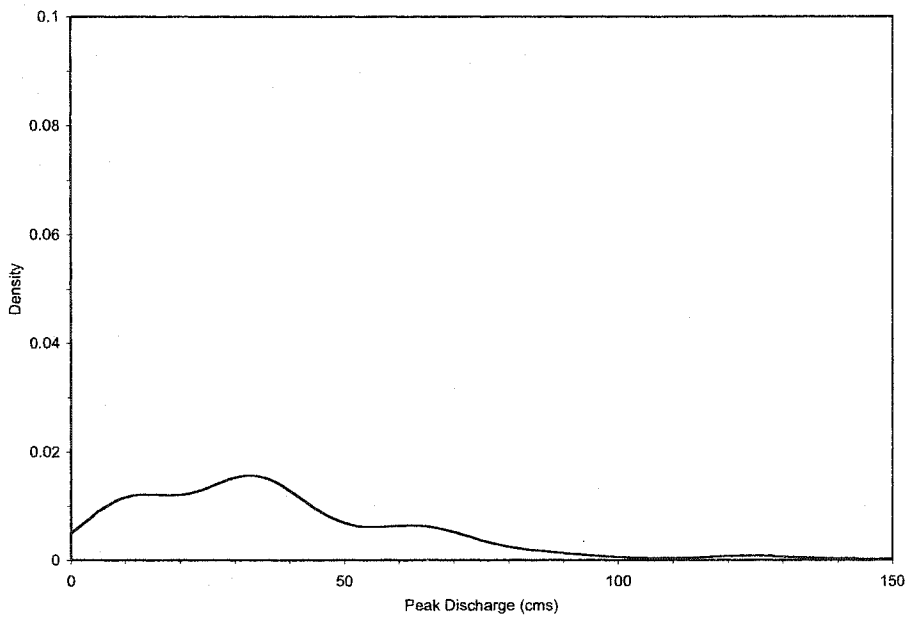


FIG. 4.45. Probability density function of the predicted magnitude of maximum discharge (Peak) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Peak 16.4 cubic meters per second (cms).

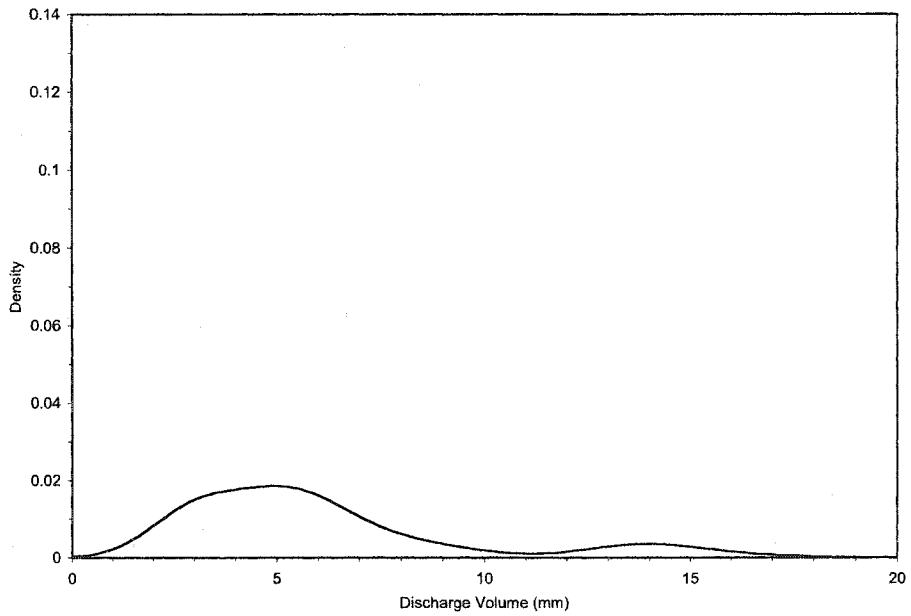


FIG. 4.46. Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 28 October 2002 case using a combined input-parameter ensemble. Observed Volume is 2.3 mm.

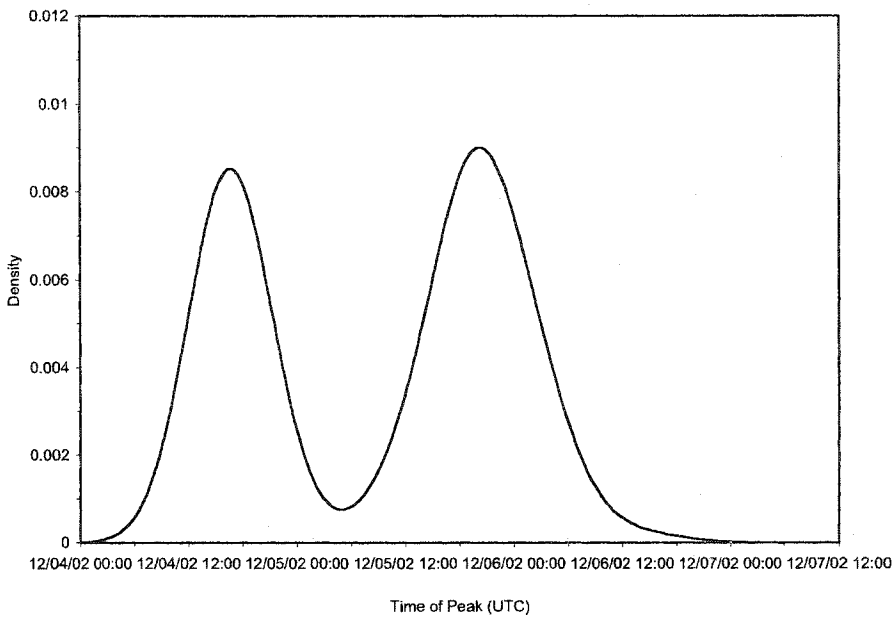


FIG. 4.47. Probability density function of the predicted time of maximum discharge (Time) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Time is 1600 UTC 05 December 2002.

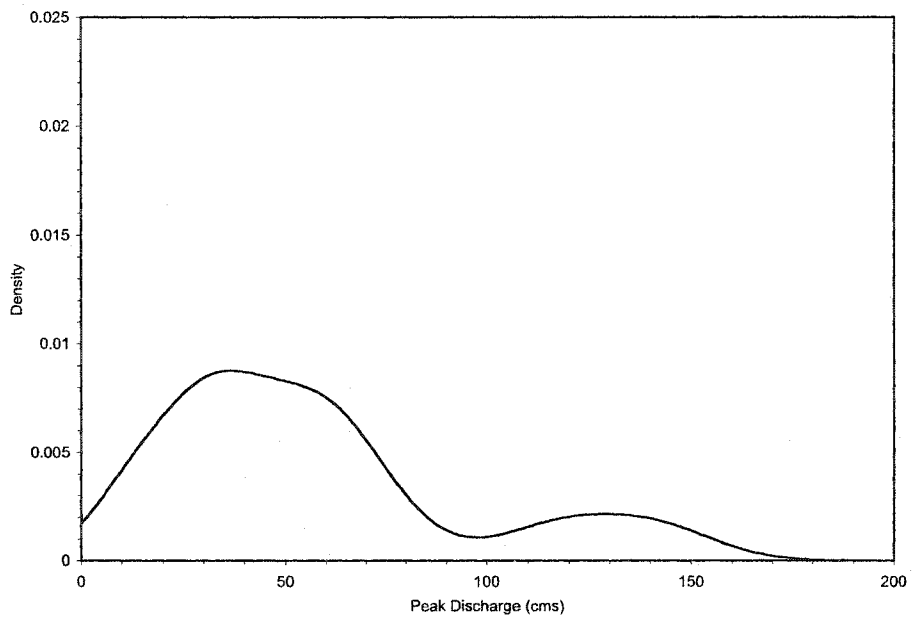


FIG. 4.48. Probability density function of the predicted magnitude of maximum discharge (Peak) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Peak 13.4 cubic meters per second (cms).

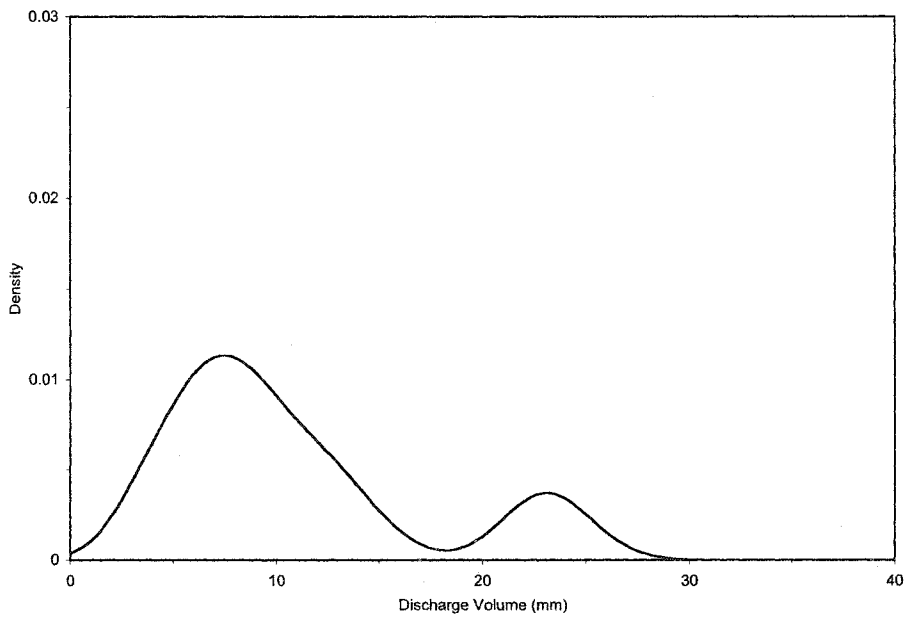


FIG. 4.49. Probability density function of the predicted time-integrated discharge normalized by the basin area (Volume) for the 03 December 2002 case using a combined input-parameter ensemble. Observed Volume is 2.5 mm.

predicted Peak and Volume pdfs (Figs. 4.48-4.49) are also falsely bimodal, but the primary density maxima correspond to observations more closely.

Simulation quantiles are computed for the combined ensembles (Figs. 4.37-4.39) for the three cool season cases. 90% simulation bounds for Time, Peak, and Volume predictions envelop observed values. There is a tendency for a majority of the members to overforecast Peak and Volume variables, but is likely an artifact of the 100% soil saturation setting alone. Hydrologic predictions, when assembled in an ensemble framework, replicate the observed streamflow behavior on the Blue River Basin for cool season, high flow events. Moreover, the predicted pdfs can now be utilized in a decision-making environment. For example, a user of the hydrologic predictions may decide that it is in their benefit to move equipment and personnel from their facility if predictions of peak discharge from the combined ensemble exceed 50 cms with a 50% probability. The medians in Fig. 4.38 show this threshold was exceeded for the 23 October 2002 and 06 December 2002 events. The developed ensembles can now be used for such purposes provided the availability of computation resources. The next section examines the number of members that are needed to replicate skill scores when the entire ensemble is used.

4.2.4. *Optimum Number of Ensemble Members*

The combined ensemble system provides an accurate pdf of possibilities expected with uncertainties in rainfall fields and model parameters. Observations of Time, Peak, and Volume all fall within the simulation limits predicted by the input-parameter ensemble for cool season cases. This system can be used for real-time application provided the number of simulations required to construct the ensemble isn't too large. Hydrologic simulations must be performed for each member. The combined ensemble for the 23 October 2002 case is comprised of 1125 members. Computational resources for this ensemble are rather demanding. It is thus worthwhile to determine the minimum number of members needed to approximate the original skill obtained with the full-blown ensemble.

Ensembles of varying sizes are created for the 23 October 2002, 28 October 2002, and 03 December 2002 cases. First, an individual member is randomly chosen from the full ensemble. Comparisons are made with observations and a *RPS* skill value is computed. This procedure is repeated again, but the number of members is increased and so on. With small ensemble sizes, there is a possibility that a given member will be an equally good or poor predictor of the hydrologic objective. For this reason, the described procedure is repeated 1000 times, and an average *RPS* value is reported for a given ensemble size.

Figures 4.50-4.52 show average *RPS* values for various ensemble sizes for all three cases. Interestingly, the curves for all cases for all hydrologic variables converge to their asymptotic *RPS* scores rather quickly. In fact, it is shown how a mere handful of members is needed to estimate the ensemble skill. Calculations are performed to determine the average number of members needed to achieve 90% of the maximum obtainable skill. All cases and all hydrologic variables are considered in this analysis. It turns out that, on average, only 10 members out of a maximum of 1125 are needed to replicate original *RPS* values. The *RPS* is computed from the entire pdf, thus it is reasonable to assume the pdfs are also accurately portrayed. This means 5%, 50%, and 95% quantiles can also be estimated from the smaller ensemble accurately. Application of the combined ensemble prediction system for real-time purposes is feasible. This study has determined that approximately 1% of the total number of available members is needed in order to estimate the predicted pdfs. Moreover, section 4.2.3 shows the predicted pdfs from the combined ensembles accurately replicate observed hydrological responses for cool season, high flow cases on the Blue River Basin.

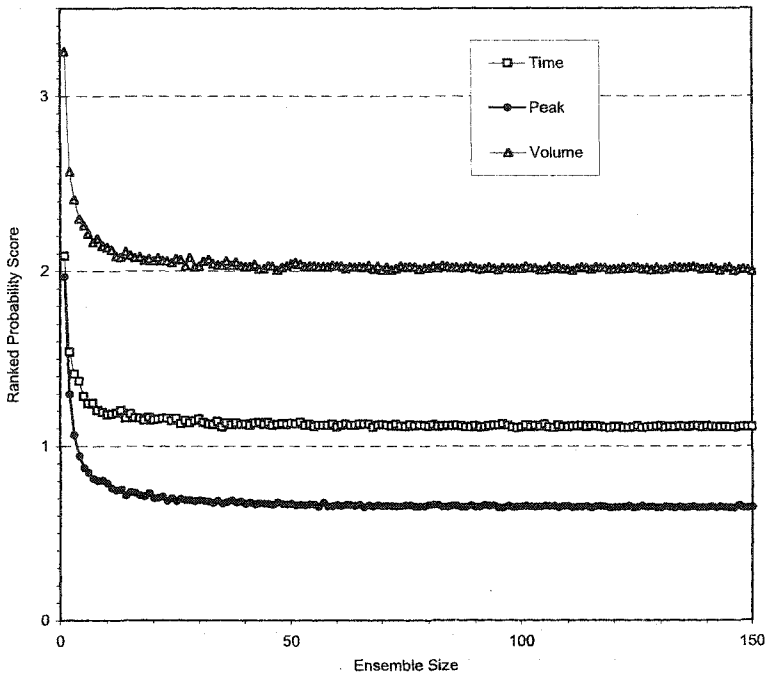


FIG. 4.50. Ranked probability score plotted as a function of the number of members comprising the ensemble for the 23 October 2002 case. The three curves correspond to predictions of time of maximum discharge (Time; open squares), magnitude of peak discharge (Peak; filled circles), and time-integrated discharge volume normalized by the basin area (Volume; open triangles).

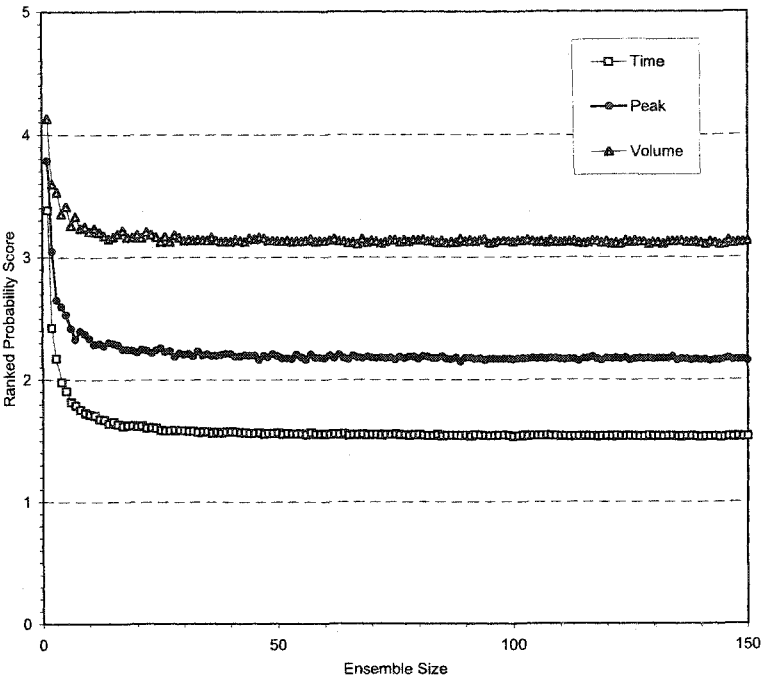


FIG. 4.51. As in Fig. 4.50 but for the 28 October 2002 case.

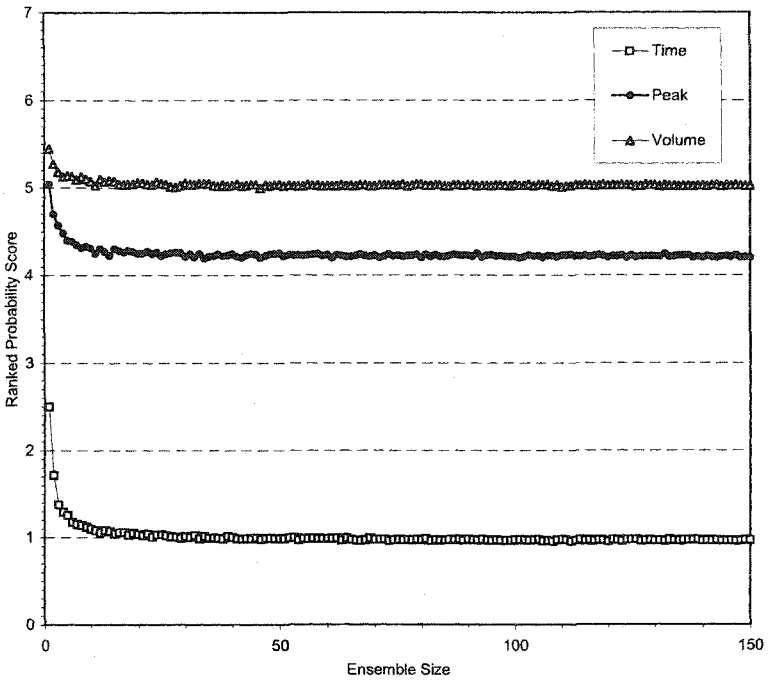


FIG. 4.52. As in Fig. 4.50 but for the 03 December 2002 case.

CHAPTER V. SUMMARY AND CONCLUSIONS

Environmental modeling involves characterizing inputs, parameters, and the physical system being modeled. The skill and uncertainty of predictions stemming from the model depend on how well these components are represented by the model. Hydrologic models transform estimates or forecasts of rainfall into streamflow predictions. In the *Vflo* ™ model, rainfall is infiltrated in the soil until rainfall rates exceed the infiltrating capabilities of the soils. Infiltration rates are determined by saturated hydraulic conductivity and initial saturation of the soils. These parameters are derived from soil classification data sets and are prone to uncertainties. After the time of ponding has been reached, excess rainfall is partitioned into overland flow. The velocity at which this overland flow reaches the channel is dictated by the Manning roughness parameter, derived from land use/cover data, in the kinematic wave equation. Finally, water is routed downstream through defined channels and the depths are converted to discharge values using rating curves. Uncertainties in streamflow predictions are a result of uncertainties in the rainfall estimates, model parameters, and in some cases the model structure. The study undertaken evaluates characteristics of uncertainty in each of these areas using an ensemble approach.

The first science question addressed herein pertains to the lack of a current systematic methodology to evaluate the skill of and monitor the improvements of

rainfall estimates (Ciach and Krajewski 1999a,b; Ciach et al. 2000). A unique approach is undertaken that places judgment on rainfall estimates not by their agreement with rain gauge measurements, but rather by the skill when the QPEs are input to a hydrologic model. Justification for improvements to quantitative precipitation estimates (QPEs) is often posed in the context of a need for improved hydrologic predictions. A methodology is devised to quantify the skill of different QPE inputs at the relevant basin scale. Application of this evaluation methodology necessarily requires that each model input is treated objectively. Specifically, a judicious selection of model parametric values can give the impression that a particular model input results in skillful hydrologic predictions. However, such a conclusion can only be justified for the unique parameter vector that was used. It is realized in environmental modeling that there may not be a unique parameter vector that produces the most skillful predictions for all situations. The presence of uncertainty in the model inputs and parameter values themselves reduces the possibility that a unique parameter set exists at all. A methodology is designed to evaluate QPE inputs by including the range of hydrologic possibilities resulting from parametric uncertainty. This ensemble approach evaluates the skill of QPE algorithms as inputs to a physics-based hydrologic model objectively and quantitatively.

Rainfall estimates are supplied to the *Vflo* ™ hydrologic model from the Quantitative Precipitation Estimation and Segregation Using Multiple Sensors

(QPE SUMS) algorithm developed at the National Severe Storms Laboratory. QPE SUMS produces a suite of rainfall products that are derived from radar, satellite, rain gauges, and combinations. Each product is input to the hydrologic model independently and hundreds of simulations are performed corresponding to each parameter combination. Probability density functions (pdfs) are used to describe the distribution of hydrologic predictions from each member. The pdf itself is then compared to observations of streamflow. A nonparametric statistic called a ranked probability score (*RPS*) is well suited for evaluating probabilistic predictions to single observations.

It is first noted that the hydrologic evaluation is performed for three cases on the Blue River Basin. Conclusions from this study may be specific to the small sample of events, characteristics of the QPE inputs, the *Vflo* ™ hydrologic model, or specific hydrologic characteristics of the Blue River Basin. Nonetheless, it is the unique methodology of evaluating QPE algorithms from the hydrologic modeling perspective that is the focus of this study. Consistencies in the results from case-to-case enable the following conclusions to be drawn about the hydrologic evaluation of QPE SUMS rainfall estimates:

1. The time of peak discharge is predicted most skillfully using inputs from the radar-based QPE SUMS product. Details in radar-based rainfall fields are important for accurate prediction of peak discharge timing.

2. Inputs from the multisensor product with mean field bias adjustment (MS-G) result in the most skillful peak discharge predictions.
3. The best predictions for the time-integrated discharge volume are accomplished using inputs from the multisensor algorithm *with no gauge adjustment.* The total volume hydrologic objective is the most revealing for QPE evaluation.
4. Three different gauge adjustment strategies to radar-based and multisensor products are evaluated. Mean field bias adjustments are found to result in superior hydrologic predictions as compared to "local" adjustment techniques. Latter techniques place more emphasis on individual rain gauge measurements, and spatial details in the original rainfall field are smoothed. These details need to be maintained for skillful hydrologic prediction.
5. Rain gauge measurements around the Blue River Basin, by themselves, do not provide an accurate depiction of the spatial variability of the rainfall field needed for accurate hydrologic prediction. Rain gauges comprising a network over a much larger region are shown to successfully adjust or calibrate QPEs.

The second science question posed in this study relates to defining the *predictability* of the hydrologic system, as requested in NRC (2001). Measuring

the predictability of a modeling system involves an assessment of prediction skill relative to a benchmark and quantification of uncertainty. Several characteristics of uncertainty are revealed in this study. First, propagation characteristics of rainfall errors to hydrologic predictions are examined. Precipitation inputs are perturbed from one-half to double their given, deterministic values. The impacts of these perturbations on streamflow predictions provides for the understanding of propagation of uncertainty through a hydrologic modeling system, the second major goal stated in NRC (1999). Ensembles are created for each rainfall perturbation by sampling the realistic parameter space for three sensitive parameters employed in the model: Manning roughness coefficient, saturated hydraulic conductivity, and initial fractional water content of the soils. The findings of this sensitivity study are summarized as follows:

1. Time of peak discharge predictions have a nonlinear response to input rainfall perturbations. The timing of peak predictions becomes early as rainfall intensities are increased.
2. Peak discharge simulations are sensitive to rainfall perturbations. While the relationship between positive rainfall perturbations and peak discharges is approximately linear, doubling the rainfall inputs results in peak discharge predictions that are increased by more than a factor of 5.

This result is not consistent with models that rely on the principle of superposition in unit hydrograph theory.

3. Predictions of time-integrated discharge volume are also sensitive to rainfall perturbations. A two-fold increase in rainfall estimates yields discharge volumes that have been amplified by more than 3 times. The relationship between uncertainty in rainfall estimates and hydrologic predictions, as measured by discharge volume, is approximately linear.
4. The uncertainty bounds associated with discharge peak and volume predictions become wider and more dispersive with larger, positive rainfall perturbations. Positive perturbations for discharge peak and volume predictions are unbound, meaning these hydrologic objectives approach infinity with larger and larger positive perturbations. On the other hand, predictions of time of peak discharge have smaller spreads with positive rainfall perturbations. Perturbations causing lighter and lighter rainfall result in time of peak discharges that occur later, asymptotically approaching positive infinity.

The next study involves identifying model structural errors or incomplete parameterizations of physical processes using a combined ensemble approach. Uncertainties may be present in model inputs, parameters, and structure. To a lesser degree, there may also be uncertainty with streamflow measurements.

Combined ensembles are constructed to explicitly account for possible hydrologic scenarios resulting from realistic errors in QPEs, sensitive model parameters, and their interactions. Observations falling outside of simulation bounds are attributed to residual uncertainty in the model structure. The following conclusions are made about structural errors in the *Vflo* ™ model:

1. Predictions of discharge peak and volume are anomalously high when the initial soil saturation parameter is set to 100%. The behavior of timing of peak discharge predictions is inconsistent, with an equal likelihood of timing predictions occurring either early or late with this parameter setting.
2. A modified version of the *Vflo* ™ model is tested to determine the effect of channel hydraulics parameterizations on streamflow predictions. A model that uses assumed trapezoidal channel geometry results in predictions that are too early and have overestimated peak discharges. The use of empirically derived rating curves better represents channel hydraulics producing more accurate predictions.
3. Streamflow predictions for a summertime, low flow case do not replicate the observed behavior, even when uncertainties in rainfall estimates and model parameters are represented in a combined ensemble. Auxiliary observations are introduced to show evidence for limited runoff

production on the Blue River Basin during summer months and its relationship to deep-layer soil moisture. The Green and Ampt submodel underestimates initial abstractions, possibly a result of enhanced flow conduits in the soil structure and interception by vegetation.

Several methods have been developed in hydrology to account for uncertainties in model parameters alone. Probabilistic predictions stemming from these ensembles are thus conditioned on perfect models, inputs, and observations of streamflow. These are unrealistic conditions. The study undertaken includes the uncertainty in model inputs and parameters in a combined ensemble. Observational uncertainties with streamflow measurements are comparatively small, and cool season, high flow cases are chosen where the model is more representative of conditions.

This is the first known attempt in hydrologic modeling to estimate the unconditioned probability density function with a given streamflow prediction. Findings regarding estimation of the total prediction uncertainty in hydrologic modeling are summarized below:

1. Simulation bounds derived from combined input-parameter ensembles encompass observations during the cool season. Hydrologic predictions, when used in an ensemble prediction system, replicate observed behavior

and provide an accurate estimate of the underlying probability density function.

2. Hydrologic predictions may now be cast in a probabilistic framework, thus enabling the computation of exceedance thresholds. This will be beneficial to decision-makers and other end-users of hydrologic predictions.
3. Probability density functions of predicted hydrologic variables can be refined by rejecting individual members (models) that do not simulate observed behavior. For example, it is suggested that members utilizing an initial soil saturation value of 100% be withheld from the combined ensembles.
4. Probabilities can be further improved by weighting individual members that have a higher or lower likelihood of replicating the observed behavior.

The ensembles used above are composed of 125 parameter combinations coupled with as many as 9 different inputs, requiring hundreds of simulations per event. Computational requirements for this large number of simulations may limit the operational relevance of a combined ensemble prediction system. A test is conducted to determine the minimum number of members needed to achieve similar skill scores as those with the full ensemble. Skill scores are computed for

varying ensemble sizes. Conclusions regarding optimal ensemble sizes are as follows:

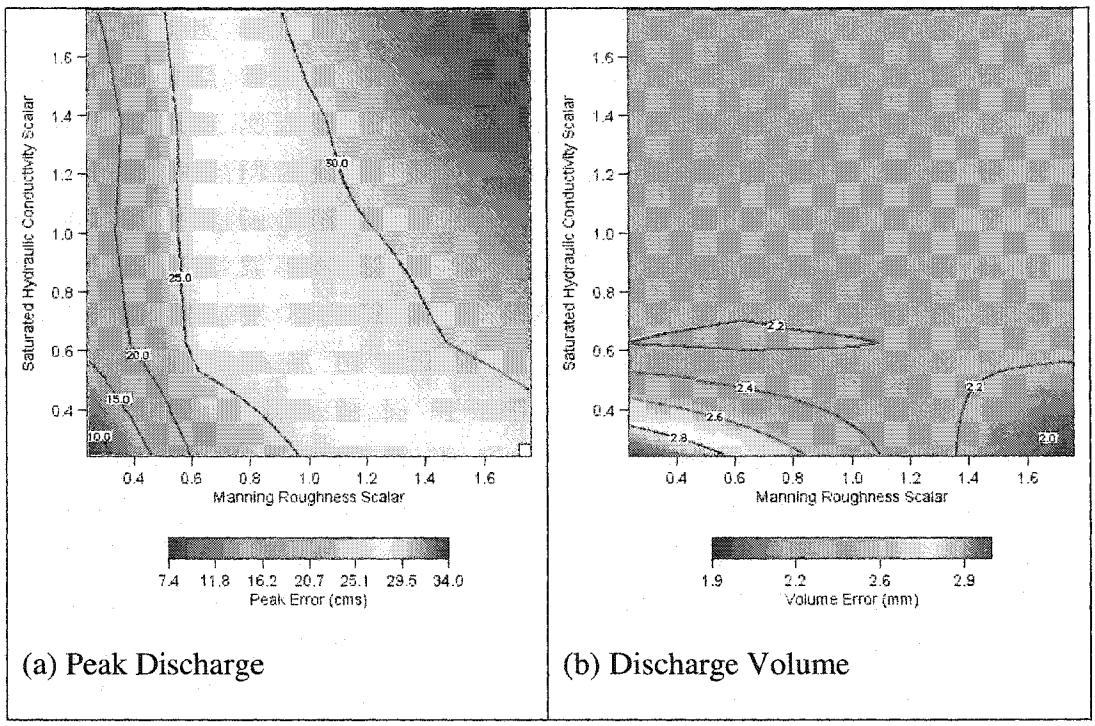
1. Ranked probability scores improve as a damped exponential with increasing ensemble sizes.
2. On average, only 10 members are needed in the combined ensemble to produce 90% of the skill obtained with the full ensemble.
3. The small number of members needed to construct a skillful combined ensemble is operationally feasible.

It is recognized that conclusions obtained for each component of this undertaking may not be generalized for all hydrologic models, other basins, or for all events on the Blue River Basin. It is the techniques employed to identify and quantify various components of uncertainty and their combined effect that make this approach to probabilistic modeling unique. Application of the developed methodologies can be used to evaluate QPE algorithms under development on any basin of interest. Moreover, the uncertainty analysis techniques can be utilized for other environmental modeling systems.

APPENDIX

The sensitivity of parameter settings to errors in peak discharge and time-integrated discharge volume is examined for the three cases used in the hydrologic evaluation (section 4.1). The error in this case is the absolute difference between simulated Peak and Volume versus observed values. Contoured error plots are produced for different settings of the saturated hydraulic conductivity and Manning roughness coefficient scalars. The error is an average computed from simulations using different settings of the initial soil saturation parameter. The initial soil saturation parameter was perturbed in the parameter adjustment studies throughout, but its sensitivity will not be shown explicitly. Parameter sensitivities are examined using inputs that result in the most skillful Volume predictions.

Figure A.1a-b shows the sensitivity of Peak and Volume predictions to parameter settings for the 23 October 2002 case. The most skillful Peak predictions (Fig. A.1a) are obtained using the smallest scalars for saturated hydraulic conductivity and Manning roughness coefficient. Minimizing these scalars will have the affect of increasing the magnitude of simulated peak discharges, suggesting either the model inputs are too low or smaller parameter scalars should have been explored. Figure A.1b shows a different response surface. In this case, there appears to be two regions of preferred parametric



(a) Peak Discharge

(b) Discharge Volume

FIG. A.1. Absolute difference between predicted and observed peak discharge (a; in cms) and time-integrated discharge volume (b; in mm) for the 23 October 2002 case. The error is an average value considering different initial soil saturation settings.

settings. Error is minimized with either the smallest available scalar of saturated hydraulic conductivity combined with the largest Manning roughness coefficient or in another region corresponding to scalars settings of approximately 0.65 for both parameters. When both plots are used in tandem, there is an indication that this latter parametric region satisfies both hydrologic objectives for this case.

The model-preferred parameter space is examined for the 28 October case (Fig. A.2a-b). Simulations of peak discharge (Fig. A.2a) are most skillful with the smallest scalar applied to saturated hydraulic conductivity maps. There doesn't appear to be much sensitivity to the Manning roughness coefficient for this case. Volume predictions (Fig. A.2b), on the other hand, are optimized with the largest scalars applied to both parameters. This result indicates model predictions generally produce too much volume and small peaks, at least for this case using MS inputs. With error minima being confined to corners and edges of the explored parameter space, future studies will need to utilize broader ranges of scalars applied to the parameter maps.

Parameteric results from the 03 December 2002 case are shown in Fig. A.3a-b. In this case, the largest scalars for both parameters are needed in order to minimize Peak and Volume errors. This case suggests either the model inputs are too large or larger parameter settings need to be explored.

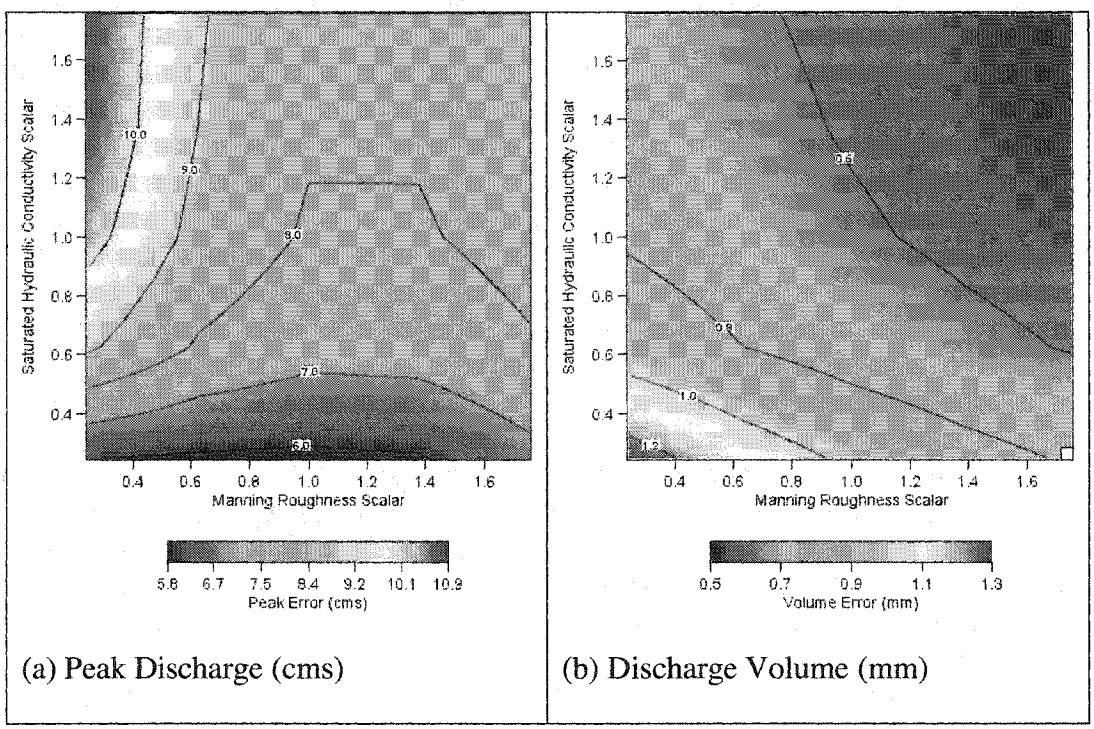


FIG. A.2. As in Fig. A.1 but for the 28 October 2002 case.

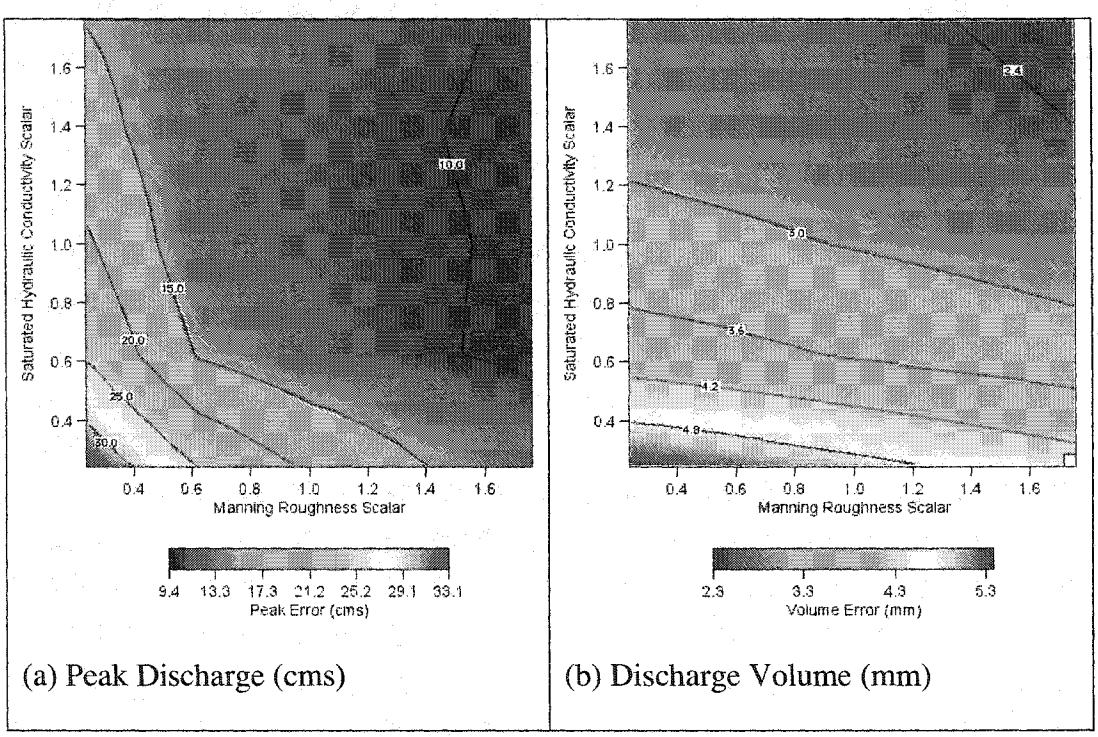


FIG. A.3. As in Fig. A.1 but for the 03 December 2002 case.

LITERATURE CITED

Abbott, M.B., J.C. Bathurst, J.A. Cunge, P.E O'Connell, and J. Rasmussen, 1986a: An introduction to European Hydrological System –Systeme Hydrologique Europeen, (SHE): History and philosophy of a physically-based distributed modeling system., *J. Hydrol.*, **87**, 45-59.

----------, J.C. Bathurst, J.A. Cunge, P.E O'Connell, and J. Rasmussen, 1986b: An introduction to European Hydrological System –Systeme Hydrologique Europeen, (SHE): Structure of a physically-based distributed modeling system, *J. Hydrol.*, **87**, 61-77.

Adler, R. F., and A. J. Negri, 1988: A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteor.*, **27**, 30-51.

----------, A. J. Negri, R. P. Keehn, and I. M. Hakkarinen, 1993: Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.*, **32**, 335-356.

Andrieu, H., and J. D. Creutin, 1995: Identification of vertical profiles of radar reflectivities for hydrological applications using an inverse method. Part I: Formulation. *J. Appl. Meteor.*, **34**, 225-239.

Arkin, P. A., 1979: The relationship between fractional coverage of high cloud and rainfall accumulations during the GATE over the B-scale array. *Mon. Wea. Rev.*, **107**, 1382-1387.

----------, and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84. *Mon. Wea. Rev.*, **115**, 51-74.

Austin, P. M., and A. Bemis, 1950: A quantitative study of the bright band in radar precipitation echoes. *J. Meteor.*, **7**, 165-171.

Barrett, E. C., and D. W. Martin, 1981: *The Use of Satellite Data in Rainfall Monitoring*. Academic Press, 340 pp.

Basara, J.B., and K.C. Crawford, 2002: Linear relationships between root-zone soil moisture and atmospheric processes in the planetary boundary layer. *J. Geophys. Res.*, **107** , (ACL 10) 1-18.

Beck, M. B., 1987: Water quality modeling: a review of the analysis of uncertainty, *Water Resour. Res.*, **23**, 1393-1442.

Beven, K. and A. Binley, 1992: The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, **6**, 279-298.

Boyles, D.P., H.V. Gupta, and S. Soorooshian, 2001: Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resour. Res.* **37**, 2749-2759.

Bolen, S., V. N. Bringi, and V. Chandrasekar, 1998: An optimal area approach to intercomparing polarimetric radar rain-rate algorithms with gauge data. *J. Atmos. Oceanic Technol.*, **15**, 605-623.

Bringi, V. N. and V. Chandrasekar, 2001: Polarimetric Doppler Weather Radar. Principles and applications. Cambridge University Press, New York, 636 pp.

Burnash, R.J.C., R.L. Ferral, and R.A. McGuire, 1973: A general stream flow simulation system - Conceptual modeling for digital computers, *Report by the Joint Federal State River Forecasts Center*, Sacramento, California.

----------, 1995: The NWS river forecast system – catchment modeling, *Computer Models of Watershed Hydrology*, ed. V.P. Singh, Water Resources Publications, Colorado, 311-366.

Carey, L. D., R. Cifelli, W. A. Petersen and S. A. Rutledge, 2000: Preliminary report on TRMM-LBA rainfall estimation using the S-POL radar. Colorado State University, *Department of Atmospheric Science,* Paper No. **697**, 19 pp.

Chow, V. T., D. R. Maidment, and L. W. Mays, 1988: *Applied Hydrology,* McGraw-Hill, Inc., McGraw-Hill Series in Water Resources and Environmental Engineering, 572 pp.

Ciach, G. J., and W. F. Krajewski, 1999a: On the estimation of radar rainfall error variance. *Advances in Water Resources*, **22,** 585-595.

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

----------, M. L. Morrissey, and W. F. Krajewski, 2000: Conditional bias in radar rainfall estimation. *J. Appl. Meteor.*, **39,** 1941-1946.

Crum, T., and R. Alberty, 1993: The WSR-88D and the WSR-88D Operational Support Facility. *Bull. Amer. Meteor. Soc.*, **74**, 1669-1687.

CUAHSI, 2002: Science Based Infrastructure for Hydrologic Sciences. Description online at: http://www.cuahsi.org/

Dotzek, N., and K. D. Beheng, 2001: The influence of deep convective motions on the variability of Z-R relations. *Atmos. Res.*, **59–60**, 15–39.

Doviak, R. J., and D. S. Zrnic, 1993: Doppler Radar and Weather Observations. 2nd Ed. Academic Press, San Diego, 562 pp.

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. Meteorol. Soc.*, **81**, 2665-2680.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125,** 2427-2459.

Duan, Q., S. Sorooshian, V.K. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015-1031.

----------, S. Sorooshian, V.K. Gupta, 1994: Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.*, **158**, 265-284.

Entekahbi, D. and co-authors, 2002: *Report of a Workshop on Predictability and Limits to prediction in Hydrologic Systems.* Committee on Hydrologic Science, National Research Council, National Academy Press, ISBN 0-309-08347-8. pp. 118.

Fabry, F., G. L. Austin, and D. Tees, 1992: The accuracy of rainfall estimates by

radar as a function of range. *Quart. J. Roy. Meteor. Soc.*, **118**, 435–453.

Federal Meteorological Handbook No. 11, 1990: Doppler radar meteorological observations, Part B, Doppler radar theory and meteorology. FCM-H11B-1990. Office of the Federal Coordinator for Meteorological Services and Supporting Research, Silver Spring, MD, 208 pp.

Frank, E., M. Borga, and E. N. Anagnostou, 1999: Hydrological modeling of mountainous basins using radar rainfall data. *Preprints. 29th Intl. Conf. On Radar Meteor.*, Amer. Meteor. Soc., Montreal, Quebec, Canada, 717-720.

Freer, J., K. Beven, and B. Ambroise, 1996: Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Res. Research*, **32**, 2161-2173.

Fulton, R. A., J. P. Breidenbach, D. J. Seo, and D. A. Miller, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377-395.

Gourley, J.J., J. Zhang, R.A. Maddox, C.M. Calvert, 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.

----------, 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.

----------, and C. M. Calvert, 2003: Automated detection of the bright band using WSR-88D radar data. *Wea. Forecasting*, **18**, 585-599.

----------, B. Kaney, and R.A. Maddox, 2003: Evaluating the calibrations of radars: A software approach. Preprints Thirty-First International Conf. On Radar Meteor., Seattle, WA, Amer. Meteor. Soc., 459-462.

Griffith, C. G., J. A. Augustine, and W. L. Woodley, 1981: Satellite rain estimation in the U.S. High Plains. *J. Appl. Meteor.*, **20**, 53-66.

Grody, N. C., 1991: Classification of snow cover and rain using the Special Sensor Microwave/Imager (SSM/I). *J. Geophys. Res.*, **96**, 7423-7435.

Gupta, V. K., S. Sorooshian, and P. O. Yapo, 1998: Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information. *Water Resour. Res.*, ***34***, 751-763.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312-1327.

Harrold, T. W., and P. G. Kitchingam, 1975: Measurement of surface rainfall using radar when the beam intersects the melting layer. Preprints: *16th Conf. on Radar Meteorology*, Boston, MA, Amer. Meteor. Soc., 473-478.

Holland, J. H., 1975: *Adaptation in Natural and Artifical Systems,* The University of Michigan Press, Ann Arbor, MI, 183 pp.

Hooke, R., and T. A. Jeeves, 1961: Direct search solutions of numerical and statistical problems. *J. Assoc. Comput. Mach.*, **8**, 212-229.

Hornberger, G. M. and co-authors, 2001: A Plan for a New Science Initiative on the Global Water Cycle. U.S. Global Change Research Program, Washington, D.C.

Hsu, K., X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.*, **36**, 1176–1190.

Hydrologic Engineering Center, 2000: The HEC Hydrologic Modeling System, Nov 95, 8 pp. (AD-A311 056)

Illingworth, A., and I. Caylor, 1989: Polarization radar estimates of raindrop size spectra and rainfall rates. *J. Atmos. Oceanic Technol.*, **6**, 939-949.

Joss, J., and A. Waldvogel, 1990: Precipitation measurements and hydrology. *Radar in Meteorology*. D. Atlas, Ed., Amer. Meteor. Soc., 577-606.

----------, and R. Lee, 1995: The application of radar-gauge comparisons to operational precipitation profile corrections. *J. Appl. Meteor.*, **34**, 2612-2630.

Julien, PY and B Saghafian, 1991: CASC2D Users Manual: A Two Dimensional Watershed Rainfall-Runoff Model. Civil Engr. Report, CER90-91PYJ-BS-12. Colorado State University, Fort Collins, CO.

----------, B. Saghafian, and F. Ogden, 1995: Raster-based hydrologic modeling of spatially-varied surface runoff. *Water Resources Bulletin,* **31,** 523-536.

Karnopp, D. C., 1963: Random search techniques for optimization problems. *Automatica,* **1,** 111-121.

Keesman, K. J., 1990: Set-theoretic parameter estimation using random scanning and principal component analysis, *Math. Comput. Simul.,* **32,** 535-544.

Kitchen, M., and P. M. Jackson, 1993: Weather radar performance at long range-Simulated and observed. *J. Appl. Meteor.*, **32**, 975-985.

Kuczera, G., and E. Parent, 1998: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.,* **211,** 69-85.

Kummerow, C. R., R. A. Mack, and I. M. Hakkarinen, 1989: A self-consistent approach to improve microwave rainfall rate estimation from space. *J. Appl. Meteor.*, **28**, 869-884.

----------, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. and Oceanic Technol.*, **15**, 809-817.

Leavesley, G.H., R. W. Lichty, B. M. Troutman, and L. G. Saindon, 1983: Precipitation-runoff modeling system--User's manual, U.S. Geological Survey Water Resources Investigation Report 83-4238, 207 p.

Legates, D. R. and T. L. deLiberty, 1993: Precipitation measurement biases in the United States. *Water Resour. Bull.,* **29,** 855-861.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409-418.

Liu, G., and J. A. Curry, 1992: Retrieval of precipitation from satellite microwave measurement using both emission and scattering. *J. Geophys. Res.*, **97**, 9959-9974.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130-141.

----------, 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289-307.

Marselek, J., 1981: Calibration of the tipping-bucket raingage. *J. Hydrol.,* **53,** 343–354.

Marshall, J. S., and W. M. Palmer, 1948: The distrubtion of raindrops with size. *J. Meteor.*, **5**, 165-166.

Masri, S. F., G. A. Bekey, and F. B. Safford, 1978: An adaptive random search method for identification of large-scale nonlinear systems. *Proc. Fourth Symp. On Identification and System Parameter Estimation,* Amsterdam, Netherlands, Int. Federation of Automatic Control, 246-255.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953: Equations of state calculations by fast computing machines, *J. Chem. Phys.*, **21,** 1087-1091.

Murphy, A. H., and M. Ehrendorfer, 1987: On the relationship between the accuracy and value of forecasts in the cost-loss ratio situation. *Wea. Forecasting*, **2**, 243-251.

Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.,* **114,** 463-494.

Nedler, J. A., and R. Mead, 1965: A simplex method for function minimization. *Comput. J.,* **7,** 308-313.

Negri, A. J., R. F. Adler, and P. J. Wetzel, 1984: Rain estimation from satellites: An examination of the Griffith-Woodley technique. *J. Climate Appl. Meteor.*, **23**, 102-116.

National Research Council, 1999: Hydrologic Science Priorities for the U.S. Global Change Research Program: An Initial Assessment. National Academy Press, Washington, D.C.

----------, 2001: Grand Challenges in Environmental Sciences, National Research Council, Committee on Grand Challenges in Environmental Sciences, Oversight Commission for the Committee on Grand Challenges in Environmental Sciences, National Academy Press, Washington D.C.

Nystuen, J. A., J. R. Proni, P. G. Black, and J. C. Wilkerson, 1996: A comparison of automatic rain gauges. *J. Atmos. and Oceanic Technol.*, **13**, 62-73.

----------, 1999: Relative performance of automatic rain gauges under different rainfall conditions. *J. Atmos. and Oceanic Technol.*, **16**, 1025-1043.

Ogden, F.L, and P.Y. Julien, 1994, Runoff model sensitivity to radar rainfall resolution, *J. Hydrol.*, **158**, 1-18.

----------, H. O. Sharif, S. U. Senarath, J. A. Smith, M. L. Baeck, and J. R. Richardson, 2000: Hydrologic analysis of the Fort Collins, Colorado flash flood of 1997. *J. Hydrology*, **228**, 82-100.

Olson, W. S., 1989: Physical retrieval of rainfall rates of over ocean by multispectral microwave radiometry: Application to tropical cyclones. *J. Geophys. Res.*, **94**, 2267-2280.

Peters, J. C. and D. J. Easton, 1996: Runoff simulation using radar rainfall data. *Water Resources Bulletin*, **32**, 753-760.

Pronzato, L., E. Walter, A. Vernot, and J. F. Lebruchec, 1984: A general-purpose global optimizer: Implementation and applications. *Math. Comput. Simul.*, **26,** 412-422.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Nino. *Mon. Wea. Rev.*, **110**, 534-584.

Romanowicz, R.J., K. Beven and J. A. Tawn, 1994: Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian Approach, *Statistics for the Environment (2), Water Related Issues*, ed. V. Barnett and F. Turkman, Wiley, Chichester, 297-318.

Rosenbrock, H. H., 1960: An automatic method of finding the greatest or least value of a function. *Comput. J.*, **3,** 175-184.

Ryzhkov, A. V., D. S. Zrnic, and D. Atlas, 1997: Polarimetrically tuned R(Z) relations and comparisons of radar rainfall methods. *J. Appl. Meteor.*, **36**, 340-349.

Sanchez-Diezma, R., D. Sempere-Torres, J. D. Creutin, I. Zawadzki, and G. Delrieu, 2001: Factors affecting the precision of radar measurement of rain. An assessment from an hydrological perspective. *Preprints. 30th Intl. Conf. On Radar Meteor.,* Amer. Meteor. Soc., Munich, Germany, 573-575.

Sauer, V. B., and Meyer, R. W., 1992, Determination of error in individual discharge measurements: U.S. Geological Survey, Open-File Report 92-144, p.

Schuur, T. J., A. V. Ryzhkov, D. S. Zrnic, and M. Schonhuber, 2001: Drop size distributions measured by a 2D video disdrometer: Comparison with dual-polarization radar data. *J. Appl. Meteor.*, **40**, 1019-1034.

Segerlind, L. J., 1984: *Applied finite element analysis.* 2nd Ed., John Wiley and Sons, New York, N.Y.

Seliga, T., and V. Bringi, 1976: Potential use of radar differential reflectivity measurements at orthogonal polarizations for measuring precipitation. *J. Appl. Meteor.*, **15**, 69-76.

Seo, D.-J., J. Breidenbach, R. Fulton, and D. Miller, 2000: Real-time adjustment of range-dependent biases in WSR-88D rainfall estimates due to nonuniform vertical profile of reflectivity. *J. Hydrometeor.*, **1**, 222-240.

----------, and J. P. Breidenbach, 2002: Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J. Hydrometeor.*, **3**, 93–111.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

Spear, R. C., and G. M. Hornberger, 1980: Eutrophication in Peel Inlet II: Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, **14**, 43-49.

Smith, C. J., 1986: The reduction of errors caused by bright bands in quantitative rainfall measurements made using radar. *J. Atmos. Oceanic Technol.*, **3**, 129-141.

Smith, J. A., D.-J. Seo, M. L. Baeck, and M. D. Hudlow, 1996: An intercomparison study of NEXRAD precipitation estimates. *Water Resour.*

Res., **32,** 2035–2045.

Sorooshian, S and J. A. Dracup, 1980: Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resour. Res.*, **29**, 1185-1194.

Spencer, R. W., 1986: A satellite passive 37 GHz scattering based method for measuring oceanic rain rates. *J. Climate Appl. Meteor.*, **25**, 754-766.

Stensrud, D. J., and J. M. Fritsch, 1994: Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.*, **122,** 2084-2104.

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

Stevenson, L., 1997: "Emerging Picture" of WSDDM's potential efficiency benefits base on field demonstrations. Draft Letter Report, available form Volpe National Transportation Systems Center, Washington, D.C.

Straka, J. M., D. S. Zrnic, A. V. Ryzhkov, 2000: Bulk hydrometeor classification and quantification using polarimetric radar data: synthesis of relations. *J. Appl. Meteor.*, **39**, 1341-1372.

Tanner, M. H., 1992: Tools for statistical inference: Observed data and data augmentation method. *Lecture Notes in Statistics*, 67, Springer-Verlag, New York, N.Y.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74,** 2317-2330.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting,* **8,** 378-398.

Van Straten, G., and K. J. Keesman, 1991: Uncerrtainty propagation and speculation in projective forecasts of environmental change: A lake-eutrophication example, *J. Forecasting,* **10,** 163-190.

Vicente, G. A., R. A. Scofield, and W. P. Menzel, 1998: The operational GOES infrared rainfall estimation technique. *Bull. Amer. Meteor. Soc.*, **79**, 1883-1898.

Vieux, B. E., and F. G. Moreda. 2003. Ordered Physics-Based Parameter Adjustment of a Distributed Model. Chapter 20 in *Advances in Calibration of Watershed Models,* ed. by Q. Duan, S. Sorooshian, H.V. Gupta, A.N. Rousseau, R. Turcotte, Monograph Series on Water Resources, American Geophysical Union, ISBN 0-87590-355-X, 267-281.

----------, and J. E. Vieux, 2002: Vflo™: A Real-Time Distributed Hydrologic Model. Available from Vieux and Associates, Inc., Norman, OK, 12 pp.

Warner, T. T., E. A. Brandes, J. Sun, D. N. Yates, and C. K. Mueller, 2000a: Prediction of a flash flood in complex terrain. Part I: A comparison of rainfall estimates from radar, and very short range rainfall simulations from a dynamic model and an automated algorithmic system. *J. Appl. Meteor.*, **39**, 797-814.

----------, D. N. Yates, and G. H. Leavesley, 2000b: A community hydrometeorology laboratory for fostering collaborative research by the atmospheric and hydrologic sciences. *Bull Amer. Meteor. Soc.*, **81**, 1499-1506.

Westrick, K. J., C. F. Mass, and B. A. Colle, 1999: The limitations of the WSR-88D radar network for quantitative precipitation estimation over the coastal western United States. *Bull. Amer. Meteor. Soc.*, **80**, 2289–2298.

----------, and C. F. Mass, 2001: An evaluation of a high-resolution hydrometeorological modeling system for prediction of a cool-season flood event in a coastal mountainous watershed. *J. Hydrometeor.*, **2**, 161-180.

Wigmosta, M.S., L.W. Vail, and D.P. Lettenmaier, 1994: A distributed hydrology-vegetation model for complex terrain, *Wat. Resour. Res.* **30**, 1665-1669.

Wilheit, T. T., and A. T. C. Chang, 1980: An algorithm for retrieval of ocean surface and atmospheric parameters from the observations of the Scanning Multichannel Microwave Radiometer. *Radio Sci.*, **15**, 525-544.

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

Wilson, J., and E. Brandes, 1979: Radar measurement of rainfall--A summary. *Bull. Amer. Meteor. Soc.*, **60**, 1048-1058.

Woodley, W. L., A. R. Olson, A. Herndon, and V. Wiggert, 1975: Comparison of gage and radar methods of convective rain measurement. *J. Appl. Meteor.*, **14**, 909-928.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1998: Multi-objective global optimization for hydrologic models, *J. Hydrol.*, **203**, 83-97.

Yates, D. N., T. T. Warner, G. H. Leavesley, 2000: Prediction of a flash flood in complex terrain. Part II: A comparison of flood discharge simulations using rainfall input from radar, a dynamic model and an automated algorithmic system. *J. Appl. Meteor.*, **39**, 815-825.

Young, B., A. A. Bradley, W. F. Krajewski, and A. Kruger, 2000: An evaluation study of NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. *J. Hydrometeor.*, **1**, 241-254.

Young, P. C., 1978: General theory of modeling for badly defined systems, *Modeling, Identification and Control of Environmental Systems,* C. G. Vansteenkiste Ed., North-Holland, Amsterdam, 103-135.

Zawadzki, I. I., 1975: On radar-raingauge comparison, *J. Appl. Meteor.*, **14**, 1430-1436.

Zrnic, D. S., and Ryzhkov, A. V., 1999: Polarimetry for weather surveillance radars. *Bull. Amer. Meteorol. Soc.*, **80**, 389-486.