Title
Hydrologic Verification: A Call for Action and Collaboration

Journal
Bulletin of the American Meteorological Society, 88(4)

ISSN
0003-0007

Authors
Welles, Edwin
Sorooshian, Soroosh
Carter, Gary
et al.

Publication Date
2007-04-01

DOI
10.1175/bams-88-4-503

HYDROLOGIC VERIFICATION

A Call for Action and Collaboration

BY EDWIN WELLES, SOROOSH SOROOSHIAN, GARY CARTER, AND BILLY OLSEN

A brief verification study of river forecasts suggests the need to link river forecast process improvements more closely to forecast verification results.

Verification must be an integral element of forecasting. Well-structured verification provides a means to improve forecast skill, to communicate with nonforecasters regarding resource needs, and to help forecast users optimize their decision making. Within the hydrology community however, few have focused any attention on verifying river forecasts. As a step toward encouraging hydrologists to verify their forecasts, this paper presents a verification study of National Oceanic and Atmospheric Administration/ National Weather Service (NWS) deterministic river-stage forecasts at 15 locations. The results of this study suggest that the hydrologic research and operations communities must join together to review, evaluate, and reconstruct the methods by which they update the hydrologic forecast process.

The verification results described in this paper are for river-stage forecasts issued by NWS River Forecast Centers (RFCs). The NWS RFCs sit at the center of the U.S. flood warning capability. They provide guidance to the NWS Weather Forecast Offices (WFOs), which in turn issue flood watches and warnings. The NWS RFCs coordinate with other state and federal water management agencies when they issue their forecasts to ensure dam operations, irrigation demand and the like are integrated into the forecasts. There are 13 RFCs across the country, and each one is responsible for a different set of basins. A more detailed description of NWS river-forecasting operations can be found in Stallings and Wenzel (1995), Larson et al. (1995), and Fread et al. (1995). In addition, the NWS RFCs describe their operations on their home pages, which can be found via the NWS home page (online at http://nws.noaa.gov).

THE CURRENT STATE OF HYDROLOGIC VERIFICATION. Though it is not commonly done, verifying hydrologic forecasts is not a new idea. When the National Research Council (NRC) reviewed the NWS hydrology program in 1996 (National Research Council 1996), one of their high-priority recommendations was the NWS implement a verification

AFFILIATIONS: WELLES—Systems Engineering Center, National Weather Service, NOAA, Silver Spring, Maryland; SOROOSHIAN—Department of Civil and Environmental Engineering, University of California, Irvine, California; CARTER—Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland; OLSEN—Arkansas–Red Basin River Forecast Center, National Weather Service, NOAA, Tulsa, Oklahoma
CORRESPONDING AUTHOR: Edwin Welles, Systems Engineering Center, National Weather Service, NOAA, 1325 East West Highway, Silver Spring, MD 20910
E-mail: edwin.welles@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-88-4-503

In final form 9 November 2006

program for the RFC river forecast time series, which form the basis for flood watches and warnings. More recently, the U.S. Department of Commerce Inspector General made a similar recommendation (Office of the Inspector General 2005), and independently, a recent NRC review (National Research Council 2006) of the Advanced Hydrologic Prediction Service (AHPS) made similar recommendations. To our knowledge, there is no evidence of any systematic hydrologic verification effort elsewhere outside of the United States. For example, the current edition of the World Meteorological Organization (WMO) "Guide to hydrological practices" (World Meteorological Organization 1994) does not mention verification.[1]

Although the effort is not extensive, the NWS has some hydrologic verification programs in place. At a national level, the NWS has verified flash flood warnings since 1986 and is working to develop a river flood warning program. The NWS has also verified river-stage forecasts, the forecasts upon which flood warnings are based, since 2001 at 173 of the 4,000 NWS river forecast locations and at all 4,000 locations since January 2006. Several NWS regions and RFCs have initiated their own hydrologic verification efforts with the NWS Central Region, verifying Mississippi and Missouri mainstem forecasts since 1983, and the Southern Region, implementing a verification program for all locations in that region. At the very local level the forecasters at the Arkansas–Red Basin River Forecast Center (ABRFC) developed their own verification program, and they publish their verification metrics online (available at www.srh.noaa.gov/abrfc). Each of these programs is a valuable contribution to hydrologic forecast verification, but none of them has had the benefit of a rigorous peer review, and few of the resulting metrics are used in the management of the NWS hydrology program. Even fewer of these metrics are used to drive the research aimed at improving hydrologic forecasts.

The hydrologic research community has focused the same scant attention on verification as the forecast community has. In general, the research community conducts the validation of new techniques through simulation, as opposed to verification, of forecasts. One example of hydrologists conducting research into hydrologic forecast techniques can be found at University of Washington West Wide Seasonal Hydrologic Forecast System (online at www.hydro.washington.edu). While the Web site provides easy access to the experimental forecasts, the corresponding verification metrics do not appear to be available on a routine basis. Despite the general trend, however, a few authors (Morris 1988; Franz et al. 2003; Pagano et al. 2004; Vivoni et al. 2003; Bradley et al. 2003, 2004) have initiated the conversation about hydrologic verification.

METHODS FOR THIS STUDY.

On the surface, the methods of verification are simple. The verification practitioner pairs the forecasts and observations, and then computes metrics to characterize the relationship between the two. In most cases, including this study, the sample of pairs is sorted into subsets, allowing the practitioner to focus on a specific aspect of the forecast process, for example, type of event, region, and so on. For this study the pairs were sorted into two subsets: a subset of pairs with the observed value below flood stage and a subset of pairs with the observed value above flood stage. That is, each forecast–observation pair was examined, and if the observation in the pair was above the flood stage, it was sorted into the "Above Flood Stage" group. If the observation was below the flood stage, it was sorted into the "Below Flood Stage" group. The forecasts were sorted by flood stage because these forecasts are used as the basis for the river flood warnings.

While the methods of verification are simple, interpreting the metrics is not always straightforward. One common interpretative aide is a reference forecast. In general, verifiers select a "dumb" forecast as a reference and then they compare the actual forecasts to these dumb forecasts. For this study, a persistence forecast is used as the dumb reference forecast. A persistence forecast is defined as the observation at the time the actual forecast is issued, persisted out into the future. See Fig. 1 for a graphical representation of a persistence forecast.

THE DATA USED FOR THIS STUDY.

A variety of verification metrics were computed for two sets of time series forecasts issued by different NWS RFCs. One dataset is from forecast locations in Oklahoma, while the other is from forecast points along the mainstem of the Missouri River. The Oklahoma (OK) dataset consists of 10 yr of forecasts and observations, starting on 1 April 1993 and ending on 30 November 2002, for four locations in Oklahoma. The Missouri Mainstem (MM) dataset consists of 20 yr of forecasts and observations, starting on 1 January 1983 and ending on 30 November 2002, for 11 locations along the mainstem of the Missouri River. Table 1 lists the total drainage area, the number of modeled upstream

---

[1] The new edition of the WMO "Guide to hydrological practices," will include a section on forecast verification (C. Barrett 2004, personal communication).

basins, the flood stage, and the record flood stage for each forecast location.

Forecast modeling environment. The forecast modeling environment consists of semidistributed, watershed-based (nongridded), locally calibrated model collections. For each basin, a set of models is linked together to represent the physical processes present in the basin. The collection of models used on individual basins will vary as the physical processes in the hydrologic system vary from basin to basin. To account for variations in the physical processes across a single watershed, a basin may be divided into subareas and a unique collection of models used on each subarea. Most commonly, forecasters divide basins to accommodate differences in snow accumulation and ablation resulting from large elevation differences between the top and the bottom of a basin in mountainous terrain. Water from upstream basins is routed through downstream basins with the local runoff added to the routed flow.

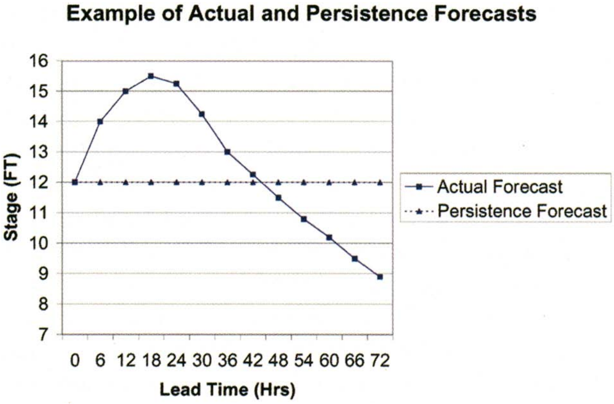

Fig. 1. An example of a theoretical forecast and a corresponding persistence forecast.

The RFC forecasters have over 30 models to choose from when setting up the models for a basin (National Weather Service 2003a). The models include runoff generation, snow accumulation, runoff routing, river routing, reservoir operations, agricultural consumptive use, and the like. The RFC forecasters configure the model collections and calibrate each model in a collection for each basin they forecast. On downstream basins, the forecasts are an accumulation of numerous model calculations. For example, from Table 1, the reader can see that the Rulo, Nebraska, location on the Missouri mainstem has 673 basins upstream of the forecast location. Therefore, there are at least 673 uniquely parameterized model collections used to forecast the Rulo flows.

Precipitation and temperature time series that drive the models are computed as areal averages for each model area. The observed precipitation input to the models is computed from gauge and Next-Generation Weather Radar (NEXRAD) rainfall estimates (Fulton et al. 1998). The use of quantitative precipitation forecasts (QPFs) varies by the forecast office. For the OK dataset, a zero QPF was used in

Table 1. Forecast point characteristics.

Location	Drainage area (mile^2)	Upstream-modeled basins	Flood stage (ft)	Record flood (ft)
Spring River at Quapaw, OK	2510	4	20.00	46.60
Illinois River at Watts, OK	635	4	13.00	25.96
Glover River at Glover, OK	315	1	16.00	29.72
Chickaskia River at Blackwell, OK	1859	2	29.00	34.38
Missouri River at South Sioux City, NE	318,559	424	30.00	30.77
Missouri River at Omaha, NE	326,759	457	29.00	40.20
Missouri River at Nebraska City NE	413,959	660	18.00	27.70
Missouri River at Rulo, NE	418,859	673	17.00	25.60
Missouri River at St. Joseph, MO	420,000	684	17.00	32.07
Missouri River at Waverly, MO	485,900	929	20.00	31.15
Missouri River at Glasgow, MO	498,900	997	25.00	39.50
Missouri River at Boonville, MO	500,700	1006	21.00	37.10
Missouri River at Jefferson City, MO	501,000	1014	23.00	38.30
Missouri River at Hermann, MO	522,500	1093	21.00	37.00
Missouri River at St. Charles, MO	524,000	1095	25.00	40.00

the early years of the data record (~1993–94); in the middle years (~1995–2000), 24 h of QPF with zero QPF beyond 24 h was used; and in the later years (~2001–03) 12 h of QPF with zero QPF beyond 12 h were used. For the MM dataset, no QPF was used until 1996 when the forecasters started using 24 h of QPF, with zero used for the QPF after 24 h. Where temperatures are used, they are computed as basin averages from stations based on distance and elevation-weighting parameters defined by the forecasters prior to forecast time. Potential evaporation (PE) is used in the OK basins, and it is computed from air temperature, dewpoint, wind speed, and solar radiation (National Weather Service 2003b). Future PE values are computed as a blend from the last observed value to the monthly climatic average.

Daily forecast process. The daily forecast process begins with quality controlling the input data and input forecasts, followed by an assessment of the model simulations, and it finishes with quality control of the output forecasts. Forecasters review all input observations and forecasts prior to using them. The method of quality control depends upon the data and the forecast office; forecasters use visual inspection of spatial trends, visual inspection of temporal trends, statistical range checking, and nearest-neighbor checks. Once the data have been reviewed and corrected as needed, the forecasters run the models and assess the simulations using visual techniques. The forecasters compare the observed and simulated stage time series in the observed period and compare the simulated model states to their expectations for the model states given the known conditions in the basins. They adjust either the model inputs, the model states, or the output of the models to cause the models to perform according to their expert knowledge of the hydrologic system and their knowledge of the abstraction of the hydrologic system into the models. Finally, they quality control the forecasts themselves by comparing their expectation of the river response given the known hydrologic conditions to the forecast river response.

Observed data arrive at the forecast offices continuously and are quality controlled as they arrive. However, the offices receive a large quantity of data shortly after 1200 UTC because most 24-h stations report at 1200 UTC, and this information must be quality controlled before it is used in the forecasts. On a busy day when numerous locations are flooded, the RFC forecasters will spend up to 6 h quality controlling the input precipitation data. The RFCs receive QPF guidance from the NWS Hydrometeorological Prediction Center (HPC) every 6 h. The forecasters may modify the HPC QPF based on current radar trends and local knowledge. Once the observations and the input forecasts have been prepared, the RFC hydrologic forecasters can begin forecasting the river conditions. Where appropriate, they contact cooperating agencies to coordinate their forecasts with the forecasts issued by the other agencies (e.g., the Army Corps of Engineers). The time requirements for issuing the forecasts vary by office and the needs of their users, but on a standard day (not a busy day), in general, all of the forecasts must be issued by 1600 UTC. Most basins are quality controlled on a daily basis; however, some forecast locations are "flood only" locations. Forecasts for these locations are not issued unless the forecasters expect the river to rise above flood stage. Forecast updates are made throughout the day as needed, depending upon user requests, the stages of the rivers, and the meteorological forecasts.

It is worth noting that the inputs and methods for hydrologic forecasting have a different structure than those used for meteorological forecasting. Numerical weather prediction models run forward into the future based upon internal model dynamics and solar forcing; therefore, any associated uncertainty is internal to the model. With the hydrologic forecasts, on the other hand, the models are driven into the future with forecast precipitation and temperature, and therefore much of the uncertainty in the forecasts is exogenous to the hydrologic modeling process. One important task for anyone conducting hydrologic verification in the future will be to analyze how the internal and external sources of error interact and limit the predictability of the hydrologic systems.

Forecast process updates over the past 20 yr. Throughout the 10- and 20-yr periods of record for the forecasts studied here, the NWS has introduced updates to the river forecasting system with the intention of improving the forecasts. The updates have varied from enhancing computing power to introducing new models and new data displays. The rainfall–runoff models were initially Antecedent Precipitation Index (API) (National Weather Service 2003c) models. The API approach has been replaced by the Sacramento Soil Moisture Accounting model (SAC-SMA) (Burnash et al. 1973) as a part of the Advanced Hydrologic Prediction Service (AHPS) initiative (McEnery et al. 2005). One objective of AHPS has been the recalibration of the SAC-SMA using longer historical datasets. In addition, the model-state updating process was enhanced with improved displays of observed and simulated variables and with improved user interfaces to support the forecasters' adjustments to the model states. The algorithm for computing the multisensor estimates of observed precipitation is also under continuous development, and it has transitioned from the original NEXRAD process (Ahnert et al. 1983, 1986; Hudlow 1988) to the current process (Seo and Breidenbach 2002; Breidenbach et al. 1999; Seo 1998). The number of precipitation gauges and the frequency of the observations has generally increased with the implementation of the Automated Surface Observing System, data collection platforms, local flood warning systems, and mesonets. The observed data quality control procedures have been enhanced through improved displays of range checking and spatial anomaly information as well as improved displays of the data itself. The precipitation forecast process has been updated because it was found that the QPF had a patchwork characteristic when the forecasts from multiple WFOs were aggregated into a single composite (National Weather Service 1999; Charba et al. 2003). The QPF is now generated at the RFC rather than WFOs.

These changes seem considerable because all aspects of the forecast process have been updated. However, it is important to realize that the fundamental forecast process has not been altered. It still consists of spatially averaged precipitation, lumped snow and runoff models, a unit hydrograph to route water in a basin, hydrologic routing techniques (as opposed to hydraulic techniques) to route water from basin to basin, and rating curves to convert computed flows to stages. In addition, the process continues to rely upon human intervention for state updating.

SUMMARY OF RESULTS.

In this study, the root-mean-square error (RMSE) is used as the single verification metric to evaluate the forecast performance for the two datasets examined. While a number of other matrices were used in the comprehensive study (Welles 2005), the conclusions drawn from the large set of metrics are the same as those seen in the RMSE. The results fall into the following three categories: those identifying success, those identifying opportunities for improvement, and those identifying a need for change. The reader will notice the much greater variability in the OK metrics compared to the MM metrics. The MM basins are much larger; therefore, the stages vary more slowly at the MM locations leading to more accurate forecasts and less variability in the verification metrics.

Results showing success. The NWS river-stage forecasts show skill and accuracy in the below–flood stage category for days 1, 2, and 3, as well as on day 1 for the above–flood stage category. The RMSE for the OK dataset below–flood stage category is plotted in Fig. 2 and for the MM dataset in Fig. 3. In Fig. 2 one can see that the RMSE stays below 2 ft for the OK dataset. In Fig. 3, it stays below 1.5 ft for the MM dataset, except in the first year. For the OK dataset, the day-3-issued forecasts show almost 1 ft less error than the day-2 persistence. The improvement over the persistence indicates that the forecasts are skillful; they are more accurate than the reference, even out to day 3. One important question with all verification is confidence. For these metrics, the sample sizes are large. Table 2 shows the minimum and maximum sample sizes for the subsets of each dataset. Though the samples are correlated both in time and space, these sample sizes are large enough to support confidence in these metrics.

The day-1 above–flood stage RMSE for both datasets is presented in Figs. 4 and 5. The breaks in the MM metrics (Fig. 5) in 1988 and 2000 indicate there were no observations above flood stage in those years. Again, the actual forecasts have a lower RMSE than the persistence. As one should expect, the above–flood stage forecasts are less accurate

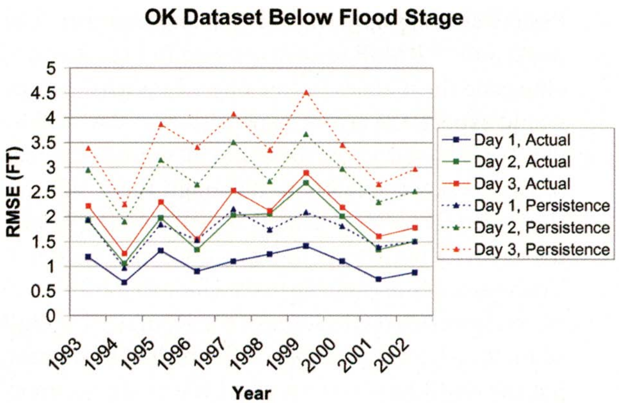

FIG. 2. The below–flood stage RMSE for the OK dataset days 1, 2, and 3.

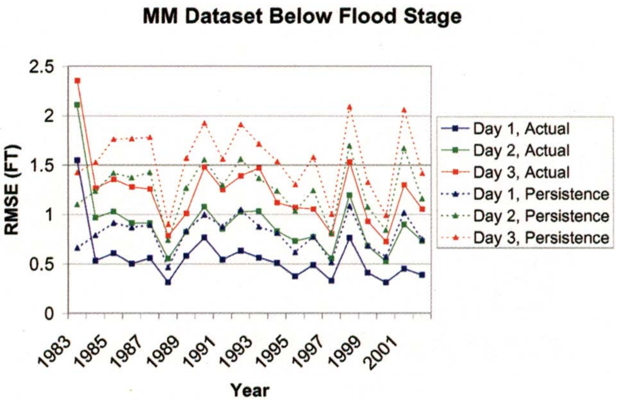

FIG. 3. The below–flood stage RMSE for the MM dataset days 1, 2, and 3.

Table 2. Summary of sample sizes for each category.

Dataset name	Above or below flood stage	Sample size range (samples)	Years in range	Min in range (samples)	Max in range (samples)
OK	Below FS	> 2,500	All years	2,100	7,000
	Above FS	0–50	1994, 1997, 1999, 2000, 2001	10	50
		> 50	1993, 1995, 1996, 1998, 2002	50	265
MM	Below FS	> 2,500	All years	2,600	3,600
	Above FS	0–100	1988, 1989, 1990, 1991, 1992, 1994, 2000, 2001, 2002	0	100
		> 100	1983, 1984, 1985, 1986, 1987, 1993, 1995, 1996, 1997, 1998, 1999	100	700

than the below–flood stage forecasts, especially for the OK dataset, which consists of small basins. The sample sizes for the above–flood stage metrics are very small for some years (see Table 2); therefore, the separation between the persistence and the actual forecasts is uncertain in those years. Nonetheless, it is still reasonable to characterize these above–flood stage, day-1 forecasts as successful, similar to the below flood stage case.

Results showing opportunities for improvement. The day-2 and -3 RMSE are also plotted in Figs. 4 and 5, alongside the metrics for the day-1 forecasts. As one would expect, the errors grow with each day, and by day 3 they have become similar in magnitude to those of the persistence reference. Though the RMSE for the issued forecasts is usually better than the persistence, this is not always the case, especially for day 3. Therefore, it is reasonable to conclude that the day-2-issued forecasts retain some of the accuracy and skill of the day-1 forecasts, especially in the OK dataset, but the day-3 forecasts provide little of the accuracy or skill seen in the day-1 forecasts. It is worth noting that Krzysztofowicz and Maranzano (2004) found a similar result on a different set of basins using an entirely different method.

Results showing a need for change. It is reasonable to expect some degree of improvement in the hydrologic forecast skill over the past 10 and 20 yr as a result of the enhancements made to the forecast process. This expectation does not appear to be met. The remainder of this paper considers the reasons why the forecast skill does not appear to have improved with time, and suggests an approach to ensuring progress in the next 10 years.

DISCUSSION. The assumption underlying the current paradigm for updating the hydrologic forecast process is that integrating improved science and computational methods into the forecast process will lead to more informative forecasts. Decisions regarding forecast process improvements have been based upon the experience of hydrologic science and forecasting experts and not upon objective verification measures. If the results from the limited set of forecast locations presented here are representative of the overall situation, it appears this approach has not worked and a new, more objective scheme is required.

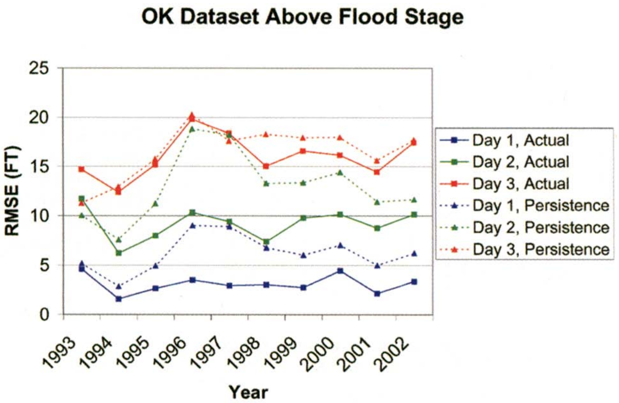

Fig. 4. The above–flood stage RMSE for the OK dataset days 1, 2, and 3.

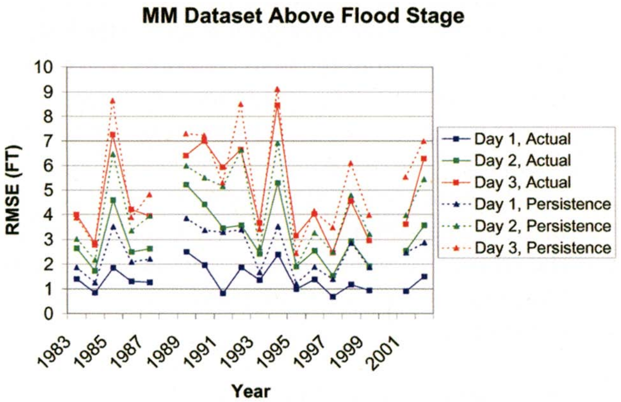

Fig. 5. The above–flood stage RMSE for the MM dataset days 1, 2, and 3.

In considering alternatives, the approach used by the meteorological forecast community serves as a good example. For instance, the development of numerical weather prediction (NWP) models is conducted by numerous, unaffiliated groups following different approaches, with the results compared through objective measures of forecast performance. In other words, the forecasts are verified, and the research is driven, not by ad hoc opinions postulated by subject matter experts, but by the actual performance of the forecasts as determined with objective measures. We suggest that hydrologists need to adopt a similar approach by embedding a comprehensive verification process into our forecast systems.

The argument in favor of driving the development of the hydrologic forecast process by objective verification measures rather than expert opinion does not rest solely on the results of the limited verification sample set provided in this paper, but it also rests upon the standards of scientific practice. Every change to the forecast process is essentially a hypothesis. In the case of the forecast process development, the hypothesis is that any changes to the forecast process will improve the system performance in some fashion. If the improvements to the hydrologic forecast process are to be considered a scientific endeavor, then any incremental change must be tested and verification is the means to do that testing. The question to be addressed is not "Should hydrologic forecasts be verified?" but "What steps must be taken to establish robust and 'universally acceptable' hydrologic forecast verification?"

In the remainder of this paper we propose two high-priority research activities to initiate the development of a comprehensive hydrologic forecast verification program. The aim will be to develop a language of collaboration between forecasters and researchers, with the goal of improving hydrologic forecasts over time. The research activities are

- define verification standards for hydrologic forecasts and use those standards to determine a comprehensive baseline description of hydrologic forecast skill; and
- quantify the sources and sinks of forecast skill, and the interaction between those sources and sinks in the hydrologic forecast process with objective measures.

The NWS started on the task of defining standards by establishing a Hydrologic Verification System Requirements (HVSR) team, which has defined the requirements for comprehensive verification system to support NWS operations. This system will identify skill and error sources, compute and display metrics, archive forecasts and observations (including model inputs), compute hindcasts, disseminate the verification results, and provide training to the forecasters. The team will submit their plan for formal publication in order to ensure this activity benefits from the collaborative input of the larger forecast community. As the science of hydrologic verification matures, the NWS will continue to benefit from collaborative input.

Verification standards and a baseline of forecast skill. Some predefined standards should serve as the foundation for a comprehensive verification strategy. While some may consider standards a limitation on "creativity," they are a means of provoking a structured discussion to define important forecast characteristics and the best techniques to evaluate those characteristics. Well-constructed standards provide a number of benefits to those of us working to enhance forecast performance. They facilitate communication by defining a common language of verification. They help forecast agencies implement operational verification procedures by providing a template for those procedures. Any set of verification standards should be published to ensure they benefit from a thorough scientific review. For an example of verification standards see the "Long range climate verification" published by the World Meteorological Organization (2002).

The NWS HVSR team is moving in the above direction. It will soon propose a set of standards that applies to both deterministic and probabilistic river forecast verification. Their proposal includes recommended metrics that define a sufficient framework to support management decisions with verification. For scientific purposes, expanded statistics are defined and referenced for users who need to understand errors and compare current and newly developed forecast methodologies. As standards are developed, both the research and operational communities can take on the substantial task of computing retrospective verification metrics in order to establish a baseline description of hydrologic forecast skill. Once the skill of the current forecasts is understood, it will be possible to identify the improvements new science brings to the forecasts.

Identify sources and sinks of forecast skill. The second priority item—quantifying the sources and sinks of forecast skill, and the interaction between those sources and sinks in the hydrologic forecast process—is the key to understanding how to improve the forecasts. If the interaction between the forecast process elements, including the forecast inputs to the hydrologic models, is not well understood, then it will not be possible to identify how changes in one element in the forecast process will change the skill of the final forecasts. Monitoring the forecast process with a well-structured verification process, which includes control forecasts as well as verification of all input forecasts, can provide some information to analyze the sources of skill in the hydrologic forecasts. For example, computing post facto forecasts with observed precipitation is one way to continuously assess the error introduced by the hydrologic model structure as opposed to the error contributed by the precipitation forecasts.

However, most of the required analyses will have to be done using hindcast methods because of the many possible forecast system configurations and the numerous and interdependent sources of errors in hydrologic forecasts. Krzysztofowicz and Maranzano (2004), Krzysztofowicz (1999), and Welles (2005) have begun this work: Krzysztofowicz and Marazano in the development of a Bayesian forecast system, and Welles in his analysis of the sources of skill on precipitation-driven headwater basins. Additional studies on snow-covered basins, on nonheadwater, on downstream basins, on basins with reservoirs, on basins with dry versus wet climates, and on basins with a variety of soil types also need to be conducted. The HVSR team will identify the suite of data NWS offices must archive and the system specifications to support this effort.

SUMMARY. In this analysis, we have identified three important points with respect to hydrologic verification. First, to date there has been little attempt to integrate verification into the hydrologic forecast process or into the research supporting the hydrologic forecast process. This shortage is reflected in the operations community at the NWS where hydrologic verification is in its infancy and in the research community where verification is neither a common topic of published papers nor included in forecast process experiments. Our analysis also indicates that there is little skill by day 3 for above–flood stage forecasts and that there has been little improvement in the forecast skill over the periods of record examined here. It is our contention that these skill characteristics are linked to the shortage of verification in hydrologic research and operations. Without conducting continuous and comprehensive verification, it is not possible to know what needs to change in order to improve the forecast skill.

THE CALL TO ACTION AND COLLABORATION. The NWS has started to lead the hydrologic community toward a robust verification tradition by taking action to standardize verification for hydrologic forecasts, document baseline forecast skill, and identify sources of skill (error) in hydrologic forecast systems. In order to support the hydrologic forecast mission, the NWS needs an in-depth understanding of the relationships between all of the components of the river forecast process and the effects of those interactions on the forecast performance. For example, the NWS must understand exactly how the implementation of a distributed model or ensemble techniques will improve forecast performance by objectively quantifying forecast error. This effort to understand the forecast process will only succeed with a strong collaboration between the hydrologic operations and research communities. Verification is the unifying theme to stimulate collaboration.

As a consequence of the absence of a rigorous verification process for operational forecasts, hydrologists have little objective information to describe the skill of their forecasts or to guide the work of improving their forecasts. Hydrologists need to take a critical and comprehensive look at their forecasts and the hydrologic forecast process. By focusing hydrologic research on forecasting, a new branch of the hydrologic discipline, a *hydrologic forecast science,*[2] will develop, and the academic community will need to update the hydrologic curriculum to include forecasting so that hydrologists can better support the complex task of providing forecasts for the nation's waterways.

ACKNOWLEDGMENTS. The second author would like to thank the National Weather Service (Grants NA87WHO582 and NA07WHO144) and the National Science Foundation STC, Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA) center in Tucson, Arizona (EAR-9876800), for their support. Also, many thanks to Bill Lawrence at the ABRFC for coding the ABRFC and Southern Region verification programs. And finally, many thanks to the reviewers who provided valuable suggestions for making this paper more readable.

REFERENCES

Ahnert, P., M. Hudlow, E. Johnson, D. Greene, and M. Rosa Dias, 1983: Proposed "on-site" precipitation processing system for NEXRAD. Preprints, *21st Conf. on Radar Meteorology,* Edmonton, AB, Canada, Amer. Meteor. Soc., 378–385.

---

[2] The term, hydrologic forecast science was first used by D. J. Seo (Office of Hydrologic Development, 2003, personal communication).

——, W. Krajewski, and E. Johnson, 1986: Kalman filter estimation of radar-rainfall field bias. Preprints, *23rd Conf. on Radar Meteorology,* Snowmass, CO, Amer. Meteor. Soc., JP33–JP37.

Bradley, A., T. Hashino, and S. S. Schwartz, 2003: Distributions-oriented verification of probability forecasts for small data samples. *Wea. Forecasting,* **18,** 903–917.

——, S. S. Schwartz, and T. Hashino, 2004: Distributions-oriented verification of ensemble streamflow predictions. *J. Hydrometeor.,* **5,** 532–545.

Breidenbach, J. P., D.-J. Seo, P. Tilles, and K. Roy, 1999: Accounting for radar beam blockage patterns in radar-derived precipitation mosaics for River Forecast Centers. Preprints, *15th Conf. on IIPS,* Dallas, TX, Amer. Meteor. Soc., 179–182.

Burnash, R., R. Ferral, and R. McGuire, 1973: A generalized streamflow simulation system: Conceptual modeling for digital computers. Joint Federal and State River Forecast Center Tech. Rep., 204 pp.

Charba, J., D. Reynolds, B. McDonald, and G. Carter, 2003: Comparative verification of recent quantitative precipitation forecasts in the National Weather Service: A simple approach for scoring forecast accuracy. *Wea. Forecasting,* **18,** 161–183.

Franz, K. J., H. C. Hartmann, S. Sorooshian, and R. Bales, 2003: Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. *J. Hydrometeor.,* **4,** 1105–1118.

Fread, D. L., and Coauthors, 1995: Modernization in the National Weather Service river and flood program. *Wea. Forecasting,* **10,** 477–484.

Fulton, R. A., J. P. Briedenbach, D. J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting,* **13,** 377–395.

Hudlow, M. D., 1988: Technological developments in real-time operational hydrologic forecasting in the United States. *J. Hydrol.,* **102,** 69–92.

Krzysztofowicz, R., 1999: Bayesian theory of probabilistic forecasting via a deterministic hydrologic model. *Water Resour. Res.,* **35,** 2739–2750.

——, and C. J. Maranzano, 2004: Hydrologic uncertainty processor for probabilistic stage transition forecasting. *J. Hydrol.,* **293,** 57–73.

Larson, L. W., and Coauthors, 1995: Operational responsibilities of the National Weather Service River and Flood Program. *Wea. Forecasting,* **10,** 465–476.

McEnery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson, 2005: NOAA's Advanced Hydrologic Prediction Service: Building pathways for better science in water forecasting. *Bull. Amer. Meteor. Soc.,* **86,** 375–385.

Morris, D., 1988: A categorical, event oriented, flood forecast verification system for National Weather Service hydrology. National Oceanographic and Atmospheric Administration/National Weather Service Tech. Memo. HYDRO 43, 74 pp.

National Research Council, 1996: *Assessment of Hydrologic and Hydrometeorological Operations and Services.* National Weather Service Modernization Committee, National Academy Press, 62 pp.

——, 2006: *Toward a New Advanced Hydrologic Prediction Service (AHPS).* Committee to Assess the National Weather Service Advanced Hydrologic Prediction Service Initiative, National Academy Press, 84 pp.

National Weather Service, 1999: Quantitative precipitation forecast process assessment. Final Rep., 72 pp.

——, 2003a: Forecast component operations. River Forecast System Manual V.3.2, 1–8.

——, 2003b: Calibration system mean areal potential evaporation computational procedure. River Forecast System Manual II.5, 1–12.

——, 2003c: The Kansas city antecedent precipitation index model. River Forecast System Manual, V.3.3, 1–8.

Office of the Inspector General, 2005: The Northeast River Forecast Center is well managed, but some improvements are needed. Inspection Rep. IPE-17259/August 2005, 42 pp.

Pagano, T., D. Garen, and S. Sorooshian, 2004: Evaluation of official western U.S. seasonal water supply outlooks, 1922–2002. *J. Hydrometeor.,* **5,** 896–909.

Seo, D.-J., 1998: Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *J. Hydrol.,* **208,** 37–52.

——, 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.

Stallings, E. A., and L. A. Wenzel, 1995: Organization of the River and Flood Program in the National Weather Service. *Wea. Forecasting,* **10,** 457–464.

Vivoni, E. R., D. Entekhabi, R. L. Bras, V. Y. Ivanov, M. P. Van Horne, C. Grassotti, and R. N. Hoffman, 2003: Quantitative flood forecasts using short-term radar nowcasting. Preprints, *17th Conf. on Hydrology,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, J4.4.

Welles, E., 2005: Verification of river stage forecasts. Ph.D. dissertation, University of Arizona, 155 pp.

World Meteorological Organization, 1994: *Guide to Hydrological Practices.* 5th ed. Publication 68, 770 pp.

——, 2002: Standardised verification system for long-range forecasts. Manual on the GDPS, Attachment II-9, WMO 485, Vol. I, 24 pp.


weatherwise
NOVEMBER/DECEMBER 2004
www.weatherwise.org
TAKING Shelter FROM THE Storm
Lessons from Hurricane Charley
Antarctic Rescue
Wellington Winds