In the NOAA’s National Weather Service (NWS), the River Forecast Centers (RFCs) produce hydrologic forecasts for a variety of applications, such as flood watches and warnings, water supply, streamflow regulation, recreational outings, and environmental impacts on ecosystems. River forecasts are either deterministic, with a single forecast value, or probabilistic, with a distribution of forecast values to account for the atmospheric, hydrologic, and anthropogenic uncertainties associated with such forecasts. RFC forecasts are issued for a wide range of space–time scales, from hours for small fast-responding basins to years for water supply forecasts in larger areas and communities.

Forecast verification is essential to monitor forecast quality over time, analyze the different sources of uncertainty and skill across the entire river forecasting process, and compare the quality of forecasts from different methodologies in order to evaluate forecast skill improvement from new science and technology. Verification should drive forecast system development and help advance the understanding of predictability. Verification information should be made available to all forecast users to guide their use of the river forecasts for better decision making.

Unfortunately, forecast verification in operational hydrology has been very limited to date, mainly due to the complexity of verifying both input forecasts (e.g., weather and climate forecasts) and hydrologic forecasts on multiple space–time scales. In the NWS, operational river forecasts have been verified by a variety of focused verification applications, but information on forecast quality has not been provided to the end users except in limited cases. Additional verification will be required to fully assess and validate existing and new forecasting components of RFC forecast systems and more clearly identify the improvements necessary to realize reliable and skillful forecasts. The need for routine and systematic hydrologic forecast verification in the NWS has been pointed out by several authors in the past (e.g., the National Research Council in 1996 and 2006, the U.S. Department of Commerce in 2005, and Welles et al. in 2007).

During the past two years, the NWS has committed substantial resources to implement routine forecast verification into the operational hydrologic forecast process, and to communicate forecast skill and uncertainty to diverse users. In 2006, an NWS advisory team published a report describing the requirements for a comprehensive river forecast verification service and a verification plan for developing and implementing such a service (www.nws.noaa.gov/oh/rfcdev/docs/Final_Verification_Report.pdf).
This paper outlines the capabilities of the NWS river forecast verification service being developed and tested, as well as the current and planned verification activities. We discuss future challenges and the need for improved collaboration between the meteorological and hydrologic research and operational communities.

ENVISIONED RIVER FORECAST VERIFICATION SERVICE. The NWS Hydrologic Services Program is developing a comprehensive river forecast verification service to evaluate the quality of delivered forecast services and the quality of all RFC forecast and guidance products. The RFC hydrologic forecast process described in Fig. 1 shows the complexity of real-time forecasting. The forecast process uses a suite of models to combine observations, analyses, and forecasts of various atmospheric and land-surface variables, including weather and climate forecasts, to represent water flow and surface processes ranging from hillslopes to tidal zones of estuaries (see documentation in NWS 2003). The verification system needs to evaluate the quality of inputs and outputs at each step along the forecast process. Verification products will be generated for a variety of users, such as scientists, forecasters, hydrology program managers, and emergency managers, as well as everyday users of hydrologic forecasts.

The river forecast verification service under development includes two components: 1) a forecast services verification or logistical component that evaluates the quality of delivered forecast services in terms of the service efficiency and forecast usability (e.g., number of forecast locations, forecast timeliness); and 2) a forecast verification component that evaluates the quality of forecasts and includes deterministic and probabilistic forecast verification on multiple space–time domains. This second component includes both diagnostic verification and real-time verification. Diagnostic verification evaluates the quality of past forecasts given certain conditions (time, forecast/observed value, event, methodology, etc.). Real-time verification evaluates and can help improve the quality of operational forecasts before the corresponding observations occur, using past observations associated with historical analogs to the real-time forecast.

In order to achieve the goals stated above, the river forecast verification service needs to provide the following capabilities:

- **Data archiving**—All data needed for a full verification service (forecasts, observations, and their attributes relative to time, service, basin, and events) must be archived regularly using standardized processes and data formats. This also includes the ability to monitor the archiving process and perform data quality control.

- **Computing verification metrics**—A variety of verification metrics needs to be computed to capture all the different aspects of forecast quality and meet the multiple needs of the users. This also requires the ability to aggregate the forecast samples to evaluate forecast quality on multiple space–time scales, as well as stratify the forecast samples using multiple conditions on forecast/observed values and other attributes.

- **Displaying verification data and metrics**—Users can examine the metrics through graphics, maps,
numerical results, and formatted reports, according to various characteristics (such as lead time), and intercompare results from different forecast sets or subsets at various space–time scales. With large amounts of data, the system will also help screen verification results and find meaningful information.

- Disseminating verification data and metrics along with documentation—The verification data and metrics, along with comprehensive documentation, will be publicly available to expose the quality and usefulness of the delivered forecasts and for collaborative verification analysis.
- Real-time access to verification metrics—The metrics will be made available in near real-time and updated regularly to communicate uncertainties both in recent forecasts and over the long term.
- Uncertainty analysis and hindcasting—Forecasters need to fully analyze their forecasts to identify sources of uncertainties and make adjustments accordingly. This requires using multiple forecast scenarios and producing hindcasts (re-forecasts) to apply the current state of the science retroactively and provide large enough samples of forecasts to verify.
- Performance measure tracking—Performance measures must be produced, reported, and tracked to show the level of success of river forecasting.

To enable collaboration on forecast verification and forecast improvement, these verification capabilities are being developed as a community verification service within NOAA’s Community Hydrologic Prediction System (CHPS); this service will be designated as CHPS Verification Service hereafter. The CHPS software infrastructure, developed using a service-oriented architecture, aims to provide the basis for improved collaboration within public, academic, and private institutions by sharing advances in science and new data and rapidly transitioning those into operational deployment.

The next section describes the status of the hydrologic verification efforts and the planned verification activities to offer such a comprehensive community verification service.

ONGOING AND PLANNED VERIFICATION ACTIVITIES. The logistical verification component, which provides a comprehensive description of service efficiency and forecast usability, includes several logistical verification measures to:

- characterize forecasts by service type, frequency, and location;
- track the number of forecasts issued by service type and location; and
- quantify the timeliness of issued forecasts, the resources for calibration and model setup, and the time required to issue each type of forecast.

New software has been developed to collect these data from the RFCs for all locations for which observed and forecast data are available. The database is currently being populated by the forecasters and will be updated with any service change. In the coming months, a set of common queries of services information will be established and prototype maps of forecast services information will be developed for dissemination to the users. The expansion of logistical measures to assess areal forecasts and evaluate additional service quality aspects is also under development.

For the river forecast verification component, development is currently focused on point forecasts that are both single-valued and probabilistic. Other agencies, including the National Centers for Environmental Prediction (NCEP), are developing and enhancing gridded forecast verification applications mainly for atmospheric forecasts, which will be integrated into the CHPS Verification Service in the future to verify gridded hydrological model inputs.

To meet the requirements of data archiving (capability 1), all RFCs store observations, deterministic forecasts used as forcing inputs, and hydrologic outputs valid at forecast points in a standardized database. However, archiving all the data needed for forecast verification and hindcasting purposes is challenging given the number of forecast points and areas, forecast types, and the number of forecast processes and methodologies that need to be evaluated. Due to current hardware and software limitations, the RFC archiving system is being redesigned as a more robust, efficient, and maintainable system for both single-valued and probabilistic forecasts.

To compute verification metrics and display data and results (capabilities 2 and 3), the operational deterministic verification capability called the Interactive Verification Program (IVP) was developed to verify deterministic forecasts of precipitation, temperature, streamflow, and stage at forecast points. Since 2001, RFCs have routinely used the IVP to verify operational stage forecasts for selected points by extracting forecast and observation data from the
RFC archive database. For probabilistic forecasts, the Ensemble Verification System (EVS) was developed to verify forcing inputs and hydrologic ensemble forecasts at forecast points, and is currently experimentally used at the RFCS.

The IVP and EVS applications will be combined into a unified verification system, called the Verification System Core, to verify operational, experimental, single-valued, and probabilistic forecasts. This system will compute a variety of verification metrics relative to bias, accuracy, skill (by comparing the forecast performance with a baseline forecast such as persistence or climatology), reliability, discrimination, and sharpness using the distributions oriented verification approach of Murphy and Winkler (1987). Such capability will enable forecasters and scientists to intercompare different forecast techniques and the performance of single-valued forecasts and probabilistic forecasts. Other system capabilities may include:

- forecast accuracy measures to assess errors related to peak values, hydrograph shape, and peak timing for specific events;
- user guidance on how to pool and stratify the forecasts to produce verification results that would be robust from large sample size and meaningful from quasihomogeneous subsets; and
- evaluation of sampling uncertainty to describe the validity of verification results via confidence intervals, which is critical for rare events.

Work is also underway to develop a real-time verification capability, which includes developing a process to select analogous forecasts from the forecast archive potentially using multiple criteria, determine specific prognostic metrics to evaluate the quality of analogous forecasts, and effectively communicate results to users. Forecasters could also correct expected forecast errors identified in the prognostic verification results.

Figure 2 gives two examples of IVP graphics from verifying deterministic streamflow forecasts. Example 2a is a plot of the stage forecast values issued at different dates and times and the corresponding observations for a flood event at a specific forecast point. Example 2b shows the root mean square error (RMSE) statistic for a range of lead times for a subset of forecasts where the corresponding observations are above flood stage; the RMSE of the forecasts is compared to the RMSE of persistence (defined as the last observed value before the forecast issuance time and repeated for all lead times in the future), to evaluate the forecast improvement over a baseline forecast. The development of the system includes verification studies of experimental and operational RFC forecasts to help enhance and validate various aspects of the system. Work is under way to publish results from these verification studies.

Other verification applications have been implemented in various parts of the NWS prior to the development of this system. The Arkansas–Red Basin RFC has developed a verification program to publish online monthly RMSE statistics for all of its daily forecast points (www.srh.noaa.gov/abrc/fctver/fctver.php). The NWS Southern Region provides categorical statistics for flood forecast verification for all four Southern Region RFCS (www.srh.noaa.gov/srh/cwwd/hsd/verification/hydro). For water supply forecasts, the NWS Western Region has developed a Web site (www.nwrfc.noaa.gov/westernwater) to archive and provide water supply forecasts for six RFCS within the western United States, as well as information on forecast quality. This application has capabilities for data visualization and computation of verification metrics. It also provides access to all forecast and verification data, as well as analysis of various climate change scenarios. We have analyzed these approaches and applied the lessons learned to develop a comprehensive and standardized verification system and a verification service user interface within the CHPS environment for all types of forecasts, in order to generate and disseminate meaningful verification products to all users.

To perform uncertainty analysis and produce hindcasts (capability 6), an ensemble hindcaster experimental prototype has been developed to retroactively apply existing and new methodologies for ensemble prediction. The EVS has been used to intercompare the performance of various sets of precipitation, temperature, and streamflow ensemble hindcasts generated with different methodologies. Hindcasting capability will be expanded to produce single-valued forecasts and evaluate the impacts of other forecasting processes (e.g., forecaster adjustments of model inputs, model states, and model outputs, postprocessing). A proper identification of sources of uncertainty needs to be developed to assess the forecast sensitivity to uncertainties in the various inputs and processes that define the entire forecasting system, and to analyze the interactions between the sources of uncertainty.

In order to efficiently develop the verification service and improve the communication of verification
ORIZED verification strategies to effectively communicate verification results to end users, as well as performance tracking measures.

The work of these teams, along with the development of the CHPS Verification Service, will be published at a later date to ensure that the NWS will benefit from the collaborative input of the larger forecast and hydrologic communities. Also, additional educational opportunities, such as short courses, workshops, and Web-based material, are being developed to help users better utilize forecast verification.

**SUMMARY AND CHALLENGES.** The NWS is making significant progress in developing a community river forecast verification service within NOAA’s CHPS to systematically evaluate and help improve river forecasts. Available NWS verification tools, such as IVP for single-valued forecasts and EVS for probabilistic forecasts, have been used to verify atmospheric and hydrologic operational and experimental forecasts on a number of basins. The CHPS Verification Service is currently being developed to include a unified verification system core and a verification service user interface to routinely produce and disseminate verification information for all types of forecasts.

To analyze the sources of forecast uncertainty, all input and output information used and produced by the river forecast system needs to be verified. Therefore, forecast verification needs to be applied and tracked across the entire NWS forecast process. Weather, climate, and water forecasts need to be evaluated on multiple space and time domains using verification metrics and parameters of hydrological relevance. It will help quantify the marginal value of improve-
ment in newly developed hydrometeorological and hydrologic observations and forecasting techniques for hydrology and water resources applications. This requires leveraging verification efforts of organizations, such as NCEP and the Earth System Research Laboratory, and adapting meteorological verification measures and practices appropriately for hydrological applications. To advance forecast verification science, collaborative research work is under way with other institutions (e.g., University of Iowa; University of California, Irvine; Iowa State; and NCAR) and with scientists involved in the Hydrologic Ensemble Prediction Experiment (HEPEX) international project.

The hydrologic and meteorological communities face similar challenges in verification science, one of which is to account for the error in the observations, which are used as the "true" values to verify the forecasts. For example, for streamflow predictions, the observational error could be accounted for by data assimilation techniques, and therefore should also be expressed in the verification metrics themselves. This is related to the development of accurate high-resolution gridded analyses and the quantification of their uncertainty to verify NWS gridded forecasts.

Another challenge concerns the selection of key verification metrics to fully describe the different aspects of the forecast quality with minimal overlapping information among the metrics. Verification results need to be interpreted by the forecasters and end users to actively aid decision making. This requires translating verification measures into scores that are physically meaningful and could influence specific decisions. Also, real-time verification promises to improve real-time forecast quality, although the selection of forecast analogs constitutes a great challenge, especially for extreme events.

This paper aims to motivate the meteorological and hydrologic research and operations communities to fully participate in the development and use of a common and modular forecast verification system to advance our understanding of the hydrologic forecasting process based on rigorous forecast verification, and to maximize the utility of forecasts. To improve hydrologic forecasting and verification science and their applications, and to deliver meaningful forecast and verification products to diverse user communities, close partnerships between scientists, forecasters, and users are needed. The NWS welcomes collaboration and support from the hydrologic research community to define new and more appropriate methods for verifying hydrologic forecasts, conduct verification studies for past forecasts, and establish benchmarks of forecast predictability.

FOR FURTHER READING


