THE SCIENCE OF NOAA'S OPERATIONAL HYDROLOGIC ENSEMBLE FORECAST SERVICE

BY JULIE DEMARGNE, LIMIN WU, SATISH K. REGONDA, JAMES D. BROWN, HAKSU LEE, MINXUE HE, DONG-JUN SEO, ROBERT HARTMAN, HENRY D. HERR, MARK FRESCH, JOHN SCHAAKE, AND YUEJIAN ZHU

HEFS extends hydrologic ensemble services from 6-hour to year-ahead forecasts and includes additional weather and climate information as well as improved quantification of major uncertainties.

As no forecast is complete without a description of its uncertainty (National Research Council of the National Academies 2006), it is necessary, for both atmospheric and hydrologic predictions, to quantify and propagate uncertainty from various sources in the forecasting system. For informed risk-based decision making, such integrated uncertainty information needs to be communicated to forecasters and users effectively. In an operational environment, ensembles are an effective means of producing uncertainty-quantified forecasts. Ensemble forecasts can be ingested in a user's downstream application (e.g., reservoir management decision support system) and used to derive probability statements about the likelihood of specific future events (e.g., probability of exceeding a flood threshold). Atmospheric ensemble forecasts have been routinely produced by operational Numerical Weather Prediction (NWP) centers for two decades. Hydrologic ensemble forecasts for long ranges have been initially based on historical observations of precipitation and temperature as plausible future inputs (e.g., Day 1985) in an attempt to account for the uncertainty at the climate time scales. Ensemble forecasts generated in this fashion were considered viable beyond 30 days where the climatic uncertainty would dominate other uncertainty sources. More recently, as the needs for risk-based management of water resources and hazards across weather and climate scales have increased, the research and operational communities have been actively working on integration of the NWP ensembles into hydrologic ensemble prediction systems and quantification of all major sources of uncertainty in such systems. In particular, the Hydrological Ensemble Prediction Experiment (HEPEX; www.hepex.org/), launched in 2004, has facilitated communications and collaborations among the atmospheric community, the hydrologic community, and the forecast users toward improving ensemble forecasts and demonstrating their utility in decision making in water management (Schaake et al. 2007b; Thielen et al. 2008; Schaake et al. 2010).

Ensemble approaches hold great potential for operational hydrologic forecasting. As demonstrated with atmospheric ensemble forecasts, the estimates of predictive uncertainty provide forecasters and users with objective guidance on the level of confidence that they may place in the forecasts. The end users can decide to take action based on their risk tolerance. Furthermore, by modeling uncertainty, hydrologic forecasters can maximize the utility of weather and climate forecasts, which are generally highly uncertain and noisy (Buizza et al. 2005). With the major uncertainties quantified and their relative importance analyzed, ensemble forecasting helps identify areas where investments in forecast systems and processes will have the greatest benefit.

Development and implementation of hydrologic ensemble prediction systems is still ongoing and hence only limited operational experience exists. A number of case studies using experimental and (pre) operational systems, however, have demonstrated their potential benefits (see, e.g., Cloke and Pappenberger 2009 and Zappa et al. 2010 for references). Recent verification studies of hydrologic ensemble forecasts or hindcasts (i.e., forecasts that are retroactively generated using a fixed forecasting system) over long time periods include Bartholmes et al. (2009), Jaun and Ahrens (2009), Renner et al. (2009), Hopson and Webster (2010), Demargne et al. (2010), Thirel et al. (2010), Van den Bergh and Roulin (2010), Addor et al. (2011), and Zappa et al. (2012) for short- to medium-range hydrologic forecasts and Kang et al. (2010), Wood et al. (2011), Fundel et al. (2012), Singla et al. (2012), and Yuan et al. (2013) for monthly to seasonal hydrologic ensembles. Objective verification analysis of ensemble forecasts or hindcasts over multiple years should improve not only the science of hydrologic ensemble forecasting but also the utility

AFFILIATIONS: Demargne—Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland, and HYDRIS Hydrologie, Saint Mathieu de Tréviers, France; Wu—Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland, and LEN Technologies, Oak Hill, Virginia; Regonda and He—Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland, and Riverside Technology, Inc., Fort Collins, Colorado; Brown—Hydrologic Solutions Limited, Southampton, United Kingdom; Lee—Office of Climate, Water, and Weather Services, National Weather Service, NOAA, Silver Spring, Maryland, and LEN Technologies, Oak Hill, Virginia; Seo—Department of Civil Engineering, The University of Texas at Arlington, Arlington, Texas; Hartman—California–Nevada River Forecast Center, National Weather Service, NOAA, Sacramento, California; Herr and Fresch—Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland; Schaake—Consultant, Annapolis, Maryland; Zhu—Environmental Modeling Center, National Centers for Environmental Prediction, National Weather Service, NOAA, College Park, Maryland
CORRESPONDING AUTHOR: Dr. Julie Demargne, 5 avenue du Grand Chêne, 34270 Saint Mathieu de Tréviers
E-mail: julie@demargne.com

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-12-00081.1

In final form 6 April 2013

of hydrologic ensemble forecast products in various downstream applications where decision support systems could be "trained" (van Andel et al. 2008).

In the National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service (NWS), an end-to-end Hydrologic Ensemble Forecast Service (HEFS) is currently being implemented as part of the Advanced Hydrologic Prediction Service (AHPS; McEnery et al. 2005) to address a variety of water information and service needs for flood risk management, water supply management, streamflow regulations, recreation planning, ecosystem management, and others (Raff et al. 2013). Such a wide range of applications requires forcing inputs and hydrologic forecasts at multiple space–time scales and for multiple forecast horizons: from minutes for flash flood predictions in fast-responding basins to years for water supply forecasts over larger areas (see examples in McEnery et al. 2005). To account for the forcing input uncertainty, the NWS River Forecast Centers (RFCs) have been using the Ensemble Streamflow Prediction (ESP) component of the National Weather Service River Forecast System (NWSRFS; National Weather Service 2012). ESP produces seasonal probabilistic forecasts of water supply based on the historical observations of precipitation and temperature (the climate being considered stationary and repeating itself) and the current hydrologic conditions (Day 1985). HEFS enhances ESP to include short-, medium-, and long-range forcing forecasts, incorporate additional weather and climate information, and better quantify the major uncertainties in hydrologic forecasting. The HEFS provides ensemble forecast and verification products and adds a major new capability to the NWS's baseline river forecasting system: the Community Hydrologic Prediction System (CHPS).

The next section presents an overview of the HEFS and its various components. In the subsequent four sections, the individual components are described in more detail and selected illustrative verification results are presented to demonstrate HEFS potential benefits. Finally, future scientific and operational challenges for improving hydrologic ensemble forecasting services are discussed.

OVERVIEW OF THE HEFS. Uncertainty in hydrologic predictions comes from many different sources: atmospheric forcing observations and predictions; initial conditions of the hydrologic model, its parameters, and structure; and streamflow regulations among other anthropogenic uncertainties (Gupta et al. 2005). The uncertainties in the atmospheric forcing inputs are typically referred to

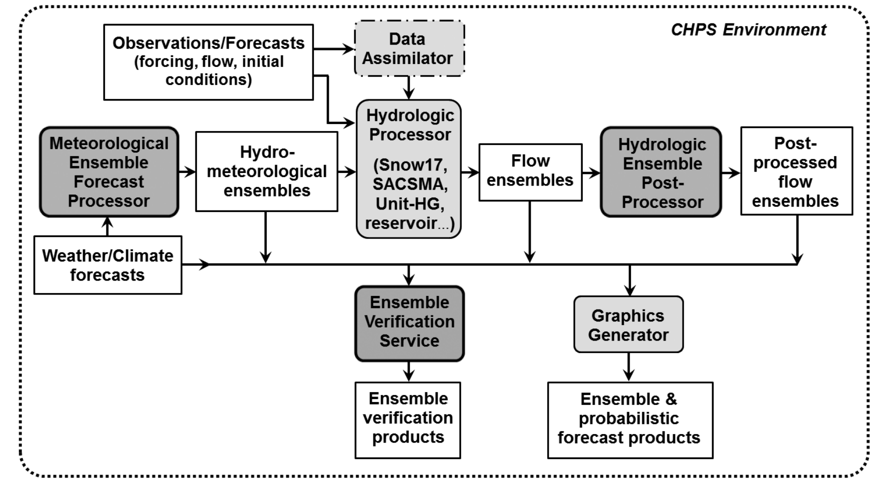

Fig. 1. Schematic view of HEFS in the NWS's CHPS, where dark gray boxes correspond to ensemble-specific components and light gray boxes are also used in operational single-valued forecasting. The current manual data assimilation procedures are not included in HEFS. The Hydrologic Processor includes a suite of hydrologic, hydraulic, and reservoir models—for example, Snow-17 model for snow ablation (Anderson 1973), the Sacramento Soil Moisture Accounting model (SAC-SMA) (Burnash 1995), and the Unit Hydrograph for flow routing.

as input uncertainty and those in all other sources as hydrologic uncertainty (Krzysztofowicz 1999). A hydrologic ensemble prediction system could either model the total uncertainty in the hydrologic output forecasts (e.g., Montanari and Grossi 2008; Coccia and Todini 2011; Weerts et al. 2011; Smith et al. 2012; Regonda et al. 2013) or explicitly account for the major sources of uncertainty, which is the primary approach of HEFS (Seo et al. 2006). As noted by Velázquez et al. (2011), hydrologic ensemble prediction systems presented in the literature often account for the input uncertainty only. Recently, however, a few systems have included various techniques to address specific hydrologic uncertainties, such as hydrologic data assimilation to reduce and model the initial condition uncertainty, Monte Carlo–based techniques to estimate model parameter uncertainty, and postprocessing and multimodel ensemble approaches for hydrologic structural uncertainty modeling (see references in Seo et al. 2006; Schaake et al. 2007b; Velázquez et al. 2011; Brown and Seo 2013; Liu et al. 2012).

A schematic view of the HEFS is given in Fig. 1 along with the information flow. For input uncertainty modeling, the Meteorological Ensemble Forecast Processor (MEFP; Schaake et al. 2007a; Wu et al. 2011) combines weather and climate forecasts from various sources to produce bias-corrected forcing (precipitation and temperature) ensembles at the space–time scales of the hydrologic models. These ensembles have coherent space–time variability among the different forcing variables and across all forecast locations. The Hydrologic Processor ingests the forcing ensembles and runs a suite of hydrologic, hydraulic, and reservoir models to produce streamflow ensembles. The data assimilation (DA) process currently consists of manual modifications of model states and parameters by the forecasters based on their expertise; therefore, it will be included in HEFS in the future when automated DA techniques are implemented. For hydrologic uncertainty modeling, the hydrologic Ensemble Postprocessor (EnsPost; Seo et al. 2006) adjusts the streamflow ensembles to reflect the total hydrologic uncertainty in a lumped manner and produce bias-corrected streamflow ensembles. Along with the above uncertainty components, the Graphics Generator and the Ensemble Verification Service (EVS; Brown et al. 2010) enable forecasters to produce uncertainty-quantified forecast and verification information that can be tailored to user needs.

Diagnostic verification of hydrologic forecasts needs to be routinely performed by scientists and operational forecasters to improve forecast quality (Welles et al. 2007) and to provide up-to-date verification information in real time to users. Such activity requires the capability of running the hydrologic ensemble forecast system in hindcasting mode to retroactively generate ensemble forecasts for multiple years using the newly developed ensemble forecasting approaches. Verification of hindcasts may be used to evaluate the benefits of new or improved ensemble forecasting approaches, analyze the various sources of uncertainty and error in the forecasting system, and guide targeted improvements of the forecasting system (Demargne et al. 2009; Renner et al. 2009; Brown et al. 2010). Hindcast datasets may also be required by operational forecasters to identify historical analog forecasts to make informed decisions in real time and by sophisticated users to calibrate decision support systems (van Andel et al. 2008). For the above reasons, the HEFS includes, for each ensemble processor, capabilities for calibration and real-time forecasting, as well as hindcasting to provide the large sample of events necessary to verify forecast probabilities with acceptable sampling uncertainty. Comprehensive evaluation of the individual HEFS components as well as the end-to-end system via multiyear hindcasting is underway (Brown 2013); illustrative examples of verification results are presented in this paper.

In the context of operational hydrologic forecasting in the NWS, the HEFS has been developed to improve upon operational single-valued forecasting and seasonal ESP forecasting while capturing user requirements, which include 1) supporting both real-time ensemble forecasting and hindcasting for large-sample verification and systematic evaluation, 2) maintaining interoperability with the single-valued forecasting system for the short range (given that single-valued forecasting is only a special case of ensemble forecasting), and 3) producing ensemble forecast information that is statistically consistent over a wide range of spatiotemporal scales. The operational hydrologic and water resources models used for both single-valued and probabilistic forecasting are simple conceptual models applied in a lumped fashion, with relatively few parameters estimated by manual calibration (a unique set of parameters being defined for all flow regimes, from low flow to flooding conditions). Expectedly, the hydrologic predictability could be limited in poorly monitored areas, with river gauges malfunctioning (e.g., during flood events) and during rapidly changing hydrometeorological conditions. Moreover, modeling reservoir regulations and diversions is challenging because of the lack of reliable information for the RFC forecasters and changes of reservoir operations to adjust to the current and forecast flow situation. Also, the estimation of historical past forcings for model calibration and hindcasting may not be consistent with real-time meteorological model inputs, owing to changes in tools (e.g., gauges versus radar for precipitation estimation) and models, as well as estimation errors. To address these data and model challenges, the RFCs have longstanding practices to apply in a subjective way manual modifications of model states and parameters for single-valued forecasting (see Raff et al. 2013 for details on RFC practices)—modifications that are not currently included in HEFS.

The initial HEFS prototype system, referred to as the Experimental Ensemble Forecast System (XEFS; www.nws.noaa.gov/oh/XEFS/), began testing at selected RFCs in 2009. The ongoing HEFS implementation is based on three software development releases to five test RFCs, from spring 2012 to fall 2013. The development phase is targeted to be completed by the end of 2013 with HEFS implementation to all 13 RFCs in 2014. The project has been accelerated by an agreement with the New York City Department of Environmental Protection, which needs these new probabilistic forecast services to more efficiently and effectively manage the water supply system for New York City (Pyke and Porter 2012).

Similarly to NWS operational single-valued hydrologic forecasting, HEFS uses CHPS, an open service-oriented architecture built on the Delft-FEWS framework (Werner et al. 2004). It facilitates incorporation of new models and tools, establishes interoperability with partners, and accelerates research to operations. CHPS is critical in supporting the NOAA Integrated Water Resources Science and Services in partnership with federal agencies [e.g., U.S. Army Corps of Engineers (USACE) and U.S. Geological Survey] that have complementary operational missions in water science, observation, prediction, and management. Also, Delft-FEWS provides an open interface to various data sources and multiple hydrologic and hydraulic forecasting models (see examples of ensemble hydrologic prediction systems based on Delft-FEWS in Werner et al. 2009, Renner et al. 2009, and Schellekens et al. 2011). Therefore, HEFS will benefit from the Delft-FEWS coupling with different hydrologic and hydraulic models, as well as enhancements made within the Delft-FEWS community for hydrologic and water resources forecast systems and services.

METEOROLOGICAL ENSEMBLE FORECAST PROCESSOR. Reliable and skillful atmospheric ensemble forecasts are necessary for hydrologic ensemble forecasting. Ensemble forecasts from NWP models are widely available from several atmospheric prediction centers. However, these ensembles are generally biased in the mean, spread, and higher moments (Buizza et al. 2005), both unconditionally and conditionally on magnitude, season, storm type, and other attributes. The conditional biases may be particularly large for heavy precipitation events that are crucial in flood forecasting (Hamill et al. 2006, 2008; Brown et al. 2012). There are several statistical techniques for estimating the conditional probability distribution of an (assumed unbiased) observed variable given a potentially biased forecast (see, e.g., references in Brown et al. 2012). These techniques vary in their assumptions about the conditional (or joint) probability distribution, the predictors used (e.g., single-valued forecast, attributes of an ensemble forecast), and the estimation of the statistical parameters (e.g., full period, seasonal, moving window, threshold dependent). Several techniques have been compared for specific variables and modeling systems (e.g., Gneiting et al. 2005; Wilks and Hamill 2007; Hamill et al. 2008). Bias correction of precipitation ensemble forecasts is particularly challenging because precipitation amount is intermittent, it depends strongly on space–time scale, and is relatively unpredictable in many cases (e.g., convective events). For hydrologic forecasting with lumped models, the gridded NWP ensembles need to be processed at the basin scale, which requires "downscaling" (described as a change of support in geostatistics) and bias correction. This downscaling includes corrections to match the climatology of the forcings used to calibrate the hydrologic model.

The MEFP aims to generate unbiased ensembles that capture the skill of the forecasts from multiple sources for individual basins while preserving the space–time properties of hydrometeorological variables (e.g., precipitation and temperature) across all basins (Schaake et al. 2007a; Wu et al. 2011). For short-range forecasts, human forecasters generally add significant value to single-valued hydrometeorological forecasts derived from raw NWP forecasts (Charba et al. 2003). Also, postprocessing studies have repeatedly demonstrated that most information from NWP medium-term ensembles comes from the ensemble mean (e.g., Hamill et al. 2004; Wilks and Hamill 2007). Therefore, the MEFP uses the single-valued forecasts modified by human forecasters for short-range forecast horizon (up to 7 days) and the ensemble mean forecasts from multiple NWP models for mid- to long range to generate seamless and calibrated hydrometeorological ensembles up to a 1-yr forecast horizon. Precipitation and temperature are processed slightly differently since precipitation is intermittent and highly skewed whereas the temperature distribution is nearly Gaussian. MEFP uses the normal quantile transform (NQT) to transform observed and forecast precipitation variables into normal variates. The precipitation part of MEFP also includes an explicit treatment of precipitation intermittency using the mixed-type bivariate meta-Gaussian model (Herr and Krzysztofowicz 2005), parametric and nonparametric modeling of the marginal probability distributions, and a parameter optimization under the continuous ranked probability score (CRPS; Hersbach 2000) and other criteria (see Wu et al. 2011 for details). For temperature, the MEFP procedure first generates ensembles of daily maximum and minimum temperatures, and then generates ensembles at subdaily time steps from the daily ensembles through a diurnal variation model. The above scheme is based on the same interpolation procedures used to calculate subdaily historical temperature time series and account for the diurnal cycle assumed in the operational calibration process (Anderson 1973).

For each hydrometeorological variable for a given basin (i.e., precipitation, maximum temperature, and minimum temperature), a specific forecast source, and a given forecast lead time, MEFP estimates the joint probability distribution of observations and single-valued forecasts based on a multiyear archive of the observations and forecasts. This calibration is performed for each day of the year by pooling historical observed–forecast pairs from a time window centered on that day in order to account for seasonality. In real time, given the current single-valued forecast, MEFP derives the conditional probability distribution of the observations, from which ensemble members are sampled. The ensemble members are generated for each individual time step, and then the Schaake shuffle (Clark et al. 2004) is applied to arrange the ensemble values according to the ranks of the historical observations. In this way, the produced ensemble time series preserve rank correlations for multiple lead times and basins across hydrometeorological variables (e.g., precipitation and temperature). The ensemble copula coupling approach (Schefzik et al. 2013) also aims to recover the space–time multivariate dependence structure from the raw ensembles instead of the historical observations. Both approaches are very attractive computationally, requiring only the computation of marginal ranks, and could be applied for any dimensionality. However, they are both limited in the number of postprocessed ensembles and equal to the number of observed historical years for the Schaake shuffle and to the number of raw ensembles for the ensemble copula coupling (making it difficult to use multiple forecast sources with different numbers of ensembles). For extreme events, if the NWP ensembles are skillful, the multivariate dependence structure should be contained in the raw ensembles (and therefore should be realistically described with the ensemble copula coupling approach). However, it may be lacking in the observation record for the Schaake shuffle approach owing to the likely lack of occurrences of similar events historically over the forecast horizon. If the NWP model output is strongly structured, parametric copula approaches might be used (as in Möller et al. 2013) to correct for any systematic errors in the ensemble's representation of the conditional dependence structure. However, such parametric procedures are very expensive computationally and could be limited in practice by the output dimensionality. We therefore suggest examining in the future alternative approaches, which use raw forecasts, observations, or some combination of the two (e.g., "analogs") for improved space–time rank structure.

In general, the forecast uncertainty and skill are time-scale dependent. Even though the forecast skill at the individual time steps may be limited, especially for long lead times, the skill of forecasts aggregated over multiple time steps is likely to be useful and needs to be exploited for hydrologic and water resources applications. Therefore, the MEFP calibration and ensemble forecasting procedures are also applied to a set of precipitation accumulations and temperature averages defined by the user across different forecast periods from the individual time steps (e.g., n-day events and x-month events up to the maximum available forecast horizon for each forecast source). The final ensemble members at the individual time steps are sequentially produced by the Schaake shuffle for the original and aggregated temporal scales according to increasing forecast skill at the individual scales and for the different forecast sources, with the highest skill having the greatest influence on the final values (see Schaake et al. 2007a for details).

MEFP has been experimentally implemented and evaluated at several RFCs using single-valued forecasts from various sources for a number of different forecast horizons. For the short-range forecast horizon, MEFP uses RFC operational single-valued forecasts as modified by the human forecasters. Depending on forecast locations, these forecasts are available from 1 to 5 forecast lead days for precipitation and up to 7 forecast lead days for temperature. Validation results were reported by 1) Schaake et al. (2007a) for precipitation and temperature for one basin in California, 2) Demargne et al. (2007) for precipitation ensembles (and corresponding streamflow ensembles) for five basins in Missouri and Oklahoma, and 3) Wu et al. (2011) for three basins in Pennsylvania, Arkansas and Missouri, and California.

Figure 2 (from Wu et al. 2011) shows the CRPS values of MEFP-generated 6-h precipitation ensembles for the first lead day for the North Fork of the American River basin (NFDC1; 875 km^2) near Sacramento, California, using three different methods. Method 1 uses an implicit treatment of precipitation intermittency (Schaake et al. 2007a); methods 2 and 3 model explicitly the precipitation intermittency, with method 3 adding parameter optimization based on the CRPS. Since, for single-valued forecasts, the CRPS collapses to the mean absolute error, the CRPS values are compared to the mean absolute error of the conditioning single-valued forecast. Figure 2 indicates that the quality of MEFP-generated precipitation ensembles has improved significantly with the explicit intermittency modeling and that the technique captures the skill in the conditioning single-valued forecast very well. In Fig. 3 (from Wu et al. 2011), the reliability diagram and the relative operating characteristic (ROC) curve for method 3 indicate that the MEFP precipitation

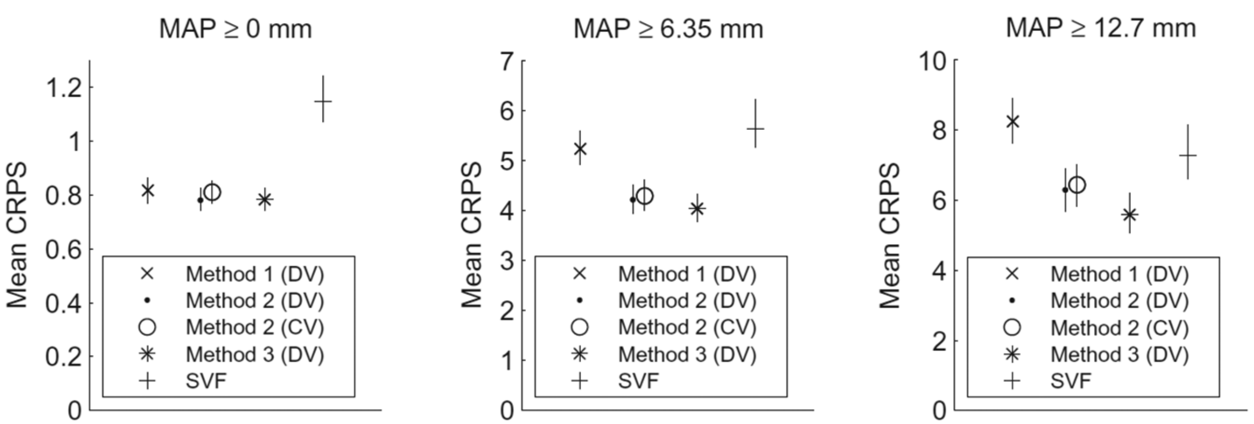

Fig. 2. Mean CRPS of ensemble hindcasts of 6-h precipitation for all four 6-h periods in day 1 for the wet season (October–May) from 2000 to 2004 in dependent validation (i.e., parameter estimation) ("DV") and independent validation (based on leave-one-year-out cross validation) ("CV"). The Mean CRPS measures the integrated squared difference between the cumulative distribution functions of a forecast and the corresponding observation, which is then averaged across a set of events. The Mean CRPS is conditioned on observed mean areal precipitation (MAP) ≥ 0, 6.35, and 12.7 mm (corresponding to nonexceedance climatological probabilities of 0.73, 0.92, and 0.96, respectively, for the wet season). The vertical bars denote the 95% confidence intervals. The results are for the NFDC1 basin with upper and lower areas combined. (From Wu et al. 2011.)

ensembles are reliable (left plot) and capture very well the discriminatory skill (right plot) in the single-valued predictions. Wu et al. (2011) show also that independent validation results (based on leave-one-year-out cross validation) are similar to dependent validation (i.e., parameter estimation) when using a historical archive of about 6 years. This suggests that the MEFP calibration with a similar or longer record length allows realizing in real-time applications the level of performance obtained in dependent validation, even if some degradation may be expected for rare events.

For medium range (up to 14 forecast lead days), the single-valued forecasts are obtained as the ensemble means from the frozen version (circa 1998) of the Global Forecast System (GFS; Hamill et al. 2006) of the National Weather Service's National Centers for Environmental Prediction (NCEP). A 30-yr reforecast archive at T62 resolution and stored on a 2.5° grid is available for the MEFP calibration. Verification results of GFS-based forcing ensembles are described in Schaake et al. (2007a) and Demargne et al. (2010) for the NFDC1 basin.

Figure 4 shows the verification results from dependent validation for the 14-day GFS-based precipitation ensembles compared to the climatology-based precipitation ensembles. All ensembles were produced at a 6-h time step from 1979 to 2005, with 45 ensemble members, using method 3, and were verified with the EVS as daily totals. The mean error is reported for the ensemble mean (commonly used as a single-valued representation of the ensemble in operational forecasting), along with the continuous ranked probability skill score (CRPSS), which describes the overall quality of the probabilistic forecast in reference to the climatology-based ensembles. At short lead times, the precipitation ensembles are relatively unbiased in the unconditional sense as evidenced by the mean error for the precipitation intermittency threshold; however, they underforecast for larger observed events. These Type-II conditional biases are common in ensemble forecasting systems since model calibration typically favors average conditions and such conditional biases are more difficult to remove with postprocessing (Brown et al. 2012). When comparing to the climatology-based ensembles for all 14 lead days, the MEFP-generated ensembles expectedly show reduced conditional biases. The quality of GFS-based precipitation decreases rapidly with increasing forecast lead time as evidenced by the increased mean

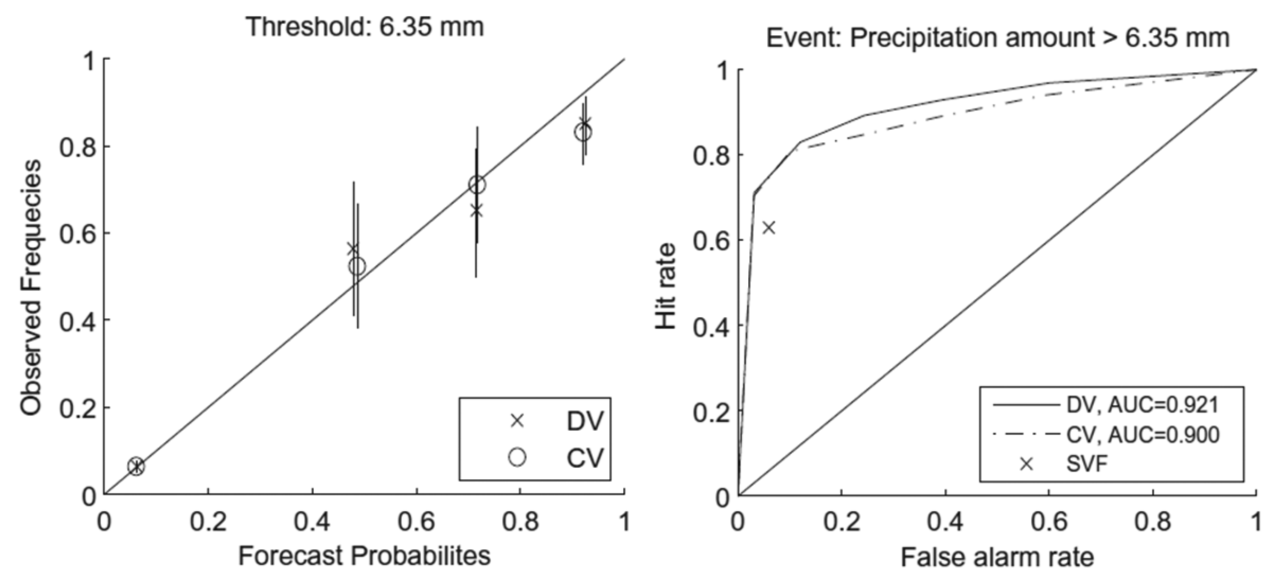

Fig. 3. (left) Reliability diagrams and (right) ROC curves relative to a threshold of 6.35 mm for MEFP-generated ensemble hindcasts of 6-h precipitation for all four 6-h periods in day 1 in DV and in leave-one-year-out CV. The reliability diagram describes the agreement between forecast probability of an event and the mean observed frequency of that event. Perfectly reliable forecasts have points that lie on the 45° diagonal line; the deviation from the diagonal line gives the conditional bias. The ROC curve plots, for a given event (e.g., flooding), the probability of detection (or hit rate) against the probability of false detection (or false alarm rate) for a range of probability levels (each one corresponding to a threshold at which a probability forecast leads to a binary decision).The false alarm rate and hit rate for the conditioning single-valued forecast are also plotted ("SVF"). The area under curve ("AUC") is defined as the area below the ROC curve and above the diagonal (corresponding to climatological forecasts), with a perfect score of 1. The vertical bars denote the 95% confidence intervals. The results are for the NFDC1 basin with upper and lower areas combined. (From Wu et al. 2011.)

error and the reduction in CRPSS. However, because of the relatively large predictability of orographic precipitation in the Sierra Nevada during the cold season in particular, GFS-based ensembles show useful skill in terms of CRPSS until 10 days.

As part of the ongoing comprehensive evaluation of HEFS ensembles, Brown (2013) analyzed verification results of GFS-based precipitation and temperature ensemble hindcasts for a 14-day forecast horizon for four pairs of headwater–downstream test basins located in California, Colorado, Kansas–Oklahoma, and Pennsylvania–New York. The GFS-based precipitation ensembles generally show skill against climatology-based ensembles for the first week but little or no skill in the second week. However, results vary significantly with basin locations (e.g., reduced precipitation predictability in the southern plains), seasons (e.g., less skill during the dry season), and magnitudes (e.g., underestimation of the probability of precipitation and, more problematically, large precipitation amounts), which underlines the need for a systematic and comprehensive evaluation of MEFP ensembles across the different RFCs.

MEFP has recently been enhanced to ingest forecast from the NCEP's latest Global Ensemble Forecast System (GEFS), which was implemented in February 2012. The new version of the GEFS uses the latest GFS model version v9.0 with an increased horizontal resolution of T254 (~55 km) for 8 days and an improved vertical resolution for all 16 days; it also includes uncertainty modeling enhancements (see Wei et al. 2008 and Hou et al. 2013, manuscript submitted to *Tellus*, for details). A new 25-yr ensemble reforecast dataset has been completed by using the configuration of the current operational GEFS and is available for public access (Hamill et al. 2013). For the longer range, the MEFP ingests single-valued forecasts from the NCEP's Climate Forecast System (CFS version 2; Saha et al. 2013), which has been operational since February 2011 and has shown some skill against climatology for hydrological ensemble forecasting (e.g., Yuan et al. 2013). MEFP constructs lagged ensemble forecasts from the single-valued CFS forecasts to estimate the ensemble mean (used as single-valued forecast to drive the MEFP statistical model) for a forecast horizon up to 9 months. MEFP requires long hindcast datasets of weather and climate forecasts from a fixed model to correct biases in the single-valued forecasts, particularly for rare events. Several studies have demonstrated that utilizing the reforecast dataset from the frozen version of a NWP model significantly improves the skill of temperature and precipitation forecasts (in particular for heavy precipitation events), as well as hydrologic forecasts (Werner et al. 2005; Hamill et al. 2006; Wilks and Hamill 2007; Hamill et al. 2008).

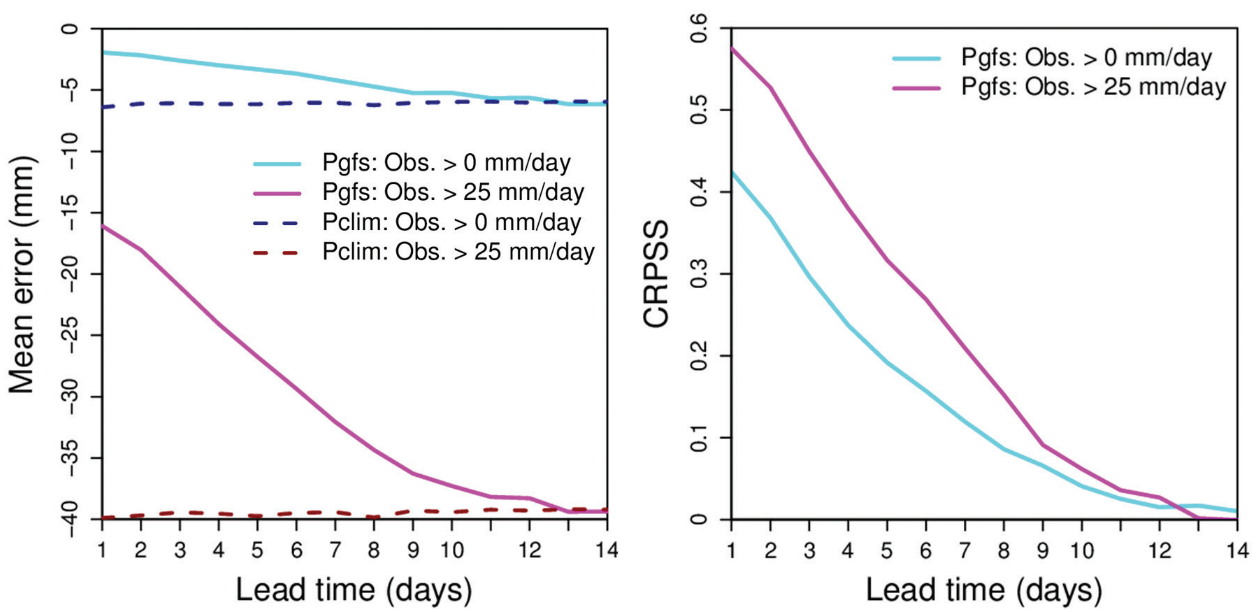

Fig. 4. (left) Mean error of the MEFP-generated ensemble means for GFS-based precipitation ensembles ("Pgfs") and climatological precipitation ensembles ("Pclim"); (right) CRPSS for the GFS-based precipitation ensembles in reference to the climatological ensembles. The CRPSS is positively oriented, with perfect skill of 1 and negative value when the forecast has worse CRPS than the reference. The results are from dependent validation for the 1979–2005 period for the NFDC1 basin with upper and lower areas combined; the threshold values of 0 and 25 mm day^{-1} correspond to nonexceedance climatological probabilities of 0.69 and 0.95, respectively.

Validation of short- to long-term ensembles for various RFC basins is underway to evaluate the expected performance of MEFP for producing seamless and skillful ensembles.

In the future, MEFP should include forecasts from other NWP models [e.g., the Short-Range Ensemble Forecast (SREF) system produced by the NCEP (Du et al. 2009)], techniques to estimate precipitation from the combination of different NWP model output variables (e.g., total column precipitable water), and additional and/or alternative postprocessing techniques, for example, to incorporate information from the ensemble spread and higher moments (Brown and Seo 2010). In the experimental Meteorological Model-Based Ensemble Forecast System (Philpott et al. 2012), three Eastern Region RFCs and a Southern Region RFC are also investigating the use of SREF and GEFS ensembles, as well as North American Ensemble Forecast System (NAEFS) ensembles, all produced and bias corrected (at the grid scale) by the NCEP (Cui et al. 2012) (experimental products available at www.erh.noaa.gov/mmefs/). Grand-ensemble datasets such as The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE; Park et al. 2008) have significant potential to capture uncertainties in the initial conditions, the model parameterization, the data assimilation technique, and the model structure through the use of atmospheric ensembles from different NWP models (e.g., Pappenberger et al. 2008; He et al. 2009, 2010). However, the use of any NWP model ensembles in hydrologic modeling requires a long reforecast dataset in order to calibrate the meteorological ensemble forecast processor as well as the hydrologic and water resources models for rare events.

HYDROLOGIC ENSEMBLE POSTPROCESSOR. Sources of hydrologic bias and uncertainty may be unknown or poorly specified in hydrologic ensemble prediction systems. Therefore, a range of statistical postprocessing techniques have been developed to account for the collective hydrologic uncertainty (Krzysztofowicz 1999; Seo et al. 2006; Coccia and Todini 2011; Brown and Seo 2013; and references therein). They aim to producing reliable (i.e., conditionally unbiased) hydrologic ensemble forecasts from single-valued forecasts or "raw" ensemble forecasts, sometimes with the aid of covariates, accounting only for the hydrologic uncertainty in the forecasts. The resulting probability distribution is described by a complete density function (e.g., Krzysztofowicz 1999; Seo et al. 2006; Montanari and Grossi 2008; Todini 2008; Bogner and Pappenberger 2011) or several thresholds of the distribution (e.g., Solomatine and Shrestha 2009; Brown and Seo 2013). Examples of postprocessing techniques for hydrologic ensemble prediction systems include error correction based on the last known forecast error (Velázquez et al. 2009), an autoregressive error correction using the most recent modeled error (Renner et al. 2009; Hopson and Webster 2010; in the latter, postprocessing is also applied to multimodel ensembles), bias correction similar to the MEFP temperature methodology for long-term ESP streamflow ensembles (Wood and Schaake 2008), a Bayesian postprocessor for ensemble streamflow forecasts (Reggiani et al. 2009), error correction for multiple temporal scales based on wavelet transformation (Bogner and Kalas 2008; Bogner and Pappenberger 2011), and a generalized linear regression model using multiple temporal scales (Zhao et al. 2011). To help establish the reliability of different statistical postprocessors and predictors under varied forecasting conditions, the HEPEX project includes an initiative to intercompare postprocessing techniques in order to develop recommendations for their operational use in hydrologic ensemble prediction systems (van Andel et al. 2012).

In the HEFS, the EnsPost (Seo et al. 2006) accounts for the collective hydrologic uncertainty in a lumped form. Since MEFP generates bias-corrected hydrometeorological ensembles that reflect the input uncertainty, EnsPost is calibrated with simulated streamflow (i.e., generated from perfect future meteorological forcings) without any manual modifications of model states and parameters. The hydrologic uncertainty is, therefore, modeled independently of forecast lead time. The postprocessed streamflow ensembles result from integration of the input and hydrologic uncertainties and hence reflect the total uncertainty. The current version of the EnsPost employs a parsimonious statistical model that combines probability matching and time series modeling. Parsimony is important to reduce data requirements and, therefore, reduce the sampling uncertainty of the estimated parameter values. The procedure adjusts each ensemble trace via recursive linear regression in the normal space (see Seo et al. 2006 for details). The regression is a first-order autoregressive model with an exogenous variable, or ARX(1,1), and uses normal-quantile-transformed historical simulation and verifying observation. The regression parameter is optimized for different seasons and flow categories, taking into account that the correlation depends greatly on flow magnitude and season. Recently, this model has been modified to better simulate temporal variability in the postprocessed streamflow ensembles by accounting for dependence in the normal space between the residual error of the model fit and the observed streamflow, as well as the serial correlation in the residual error.

EnsPost is currently applied to daily observed and forecast streamflows; after statistical postprocessing, the adjusted ensemble values are disaggregated to subdaily flows. In Seo et al. (2006) and subsequent studies for other locations, the EnsPost shows satisfactory results for short forecast horizons and for all ranges of flow. However, independent validation shows slightly degraded results in comparison to dependent validation when EnsPost parameters were estimated from a 20-yr record, mainly owing to uncertainties in the empirical cumulative distribution functions of observed and simulated flows. Seo et al. (2006) underlined that in real-time applications, when the postprocessor parameters may be regularly (e.g., annually) updated using more than 20 years of data, the performance of EnsPost would be similar or better than the obtained independent validation results. Examples of cross validation results are shown in Fig. 5 for postprocessed flow ensemble hindcasts produced with perfectly known future forcing. The daily flow ensemble hindcasts were generated for the NFDC1 basin using 38 years of observed–simulated flow records. In Fig. 5, the reliability diagram and the ROC curve relative to a threshold of 95th-percentile flow indicate good reliability (left plot) and discriminatory skill similar to the single-valued model predictions (right plot) for the first and fifth lead days. However, the current version of EnsPost is of limited utility for complex flow regulations and does not explicitly account for timing errors in the streamflow simulations (see Liu et al. 2011).

Regarding the quality of HEFS flow ensembles, examples of dependent verification results are given for raw and postprocessed flow ensemble hindcasts produced by the Hydrologic Processor and EnsPost using the GFS-based precipitation and temperature ensembles generated by MEFP. For flow hindcasting, the Hydrologic Processor is first run in simulation mode with the observed precipitation and temperature time series to generate the historical initial conditions for all hindcast dates. Based on the historical initial conditions, flow ensemble hindcasts are produced by the Hydrologic Processor for each hindcast date using the MEFP precipitation and temperature ensemble hindcasts, and then postprocessed by EnsPost. To evaluate the performance gain using MEFP and EnsPost, flow ensembles produced by the Hydrologic Processor (using the same retrospective initial conditions) from climatological forcing ensembles are used as reference forecasts.

The example verification results are given for the NFDC1 basin, for which 6-h ensemble hindcasts were

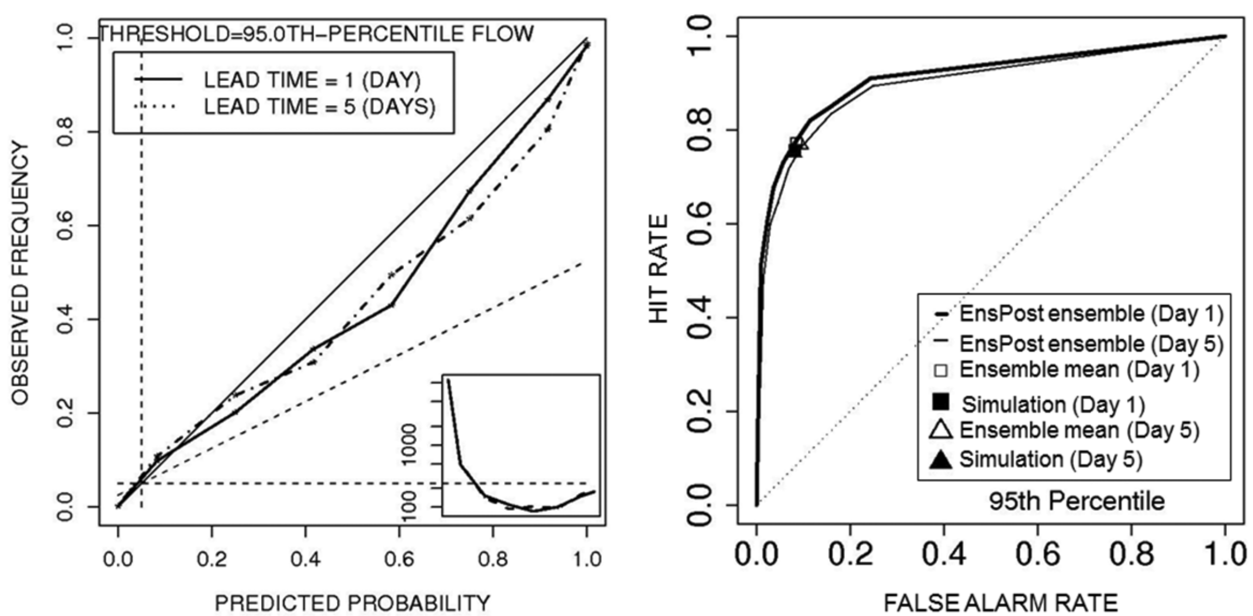

Fig. 5. (left) Reliability diagrams and (right) ROC curves relative to the 0.95 nonexceedance probability threshold for the NFDC1 basin for the 1961–99 period for EnsPost-generated flow ensemble hindcasts under the assumption of perfectly known future forcing, for days 1 and 5 in leave-one-year-out cross validation; the inset in (left) represents the histogram of predicted probability by the ensemble hindcasts.

produced from 1979 to 2005 and verified with EVS as daily average flows. The comparisons of dependent and independent validation results for MEFP and EnsPost in the previous studies (e.g., Wu et al. 2011; Seo et al. 2006) have shown their robustness. Thus, the following dependent validation results for HEFS-generated flow ensembles give a reasonable indication of the expected performance of HEFS in real-time applications, when both MEFP and EnsPost are calibrated with more than 25 years of data, even if some degradation is expected for rare events. As illustrated in Figs. 2–4 for the NFDC1 basin, MEFP precipitation ensembles perform well, particularly when compared with climatological ensembles. The marginal value of EnsPost depends largely on the magnitude of the systematic bias in the model-simulated streamflow. For the NFDC1 basin, the model simulation is of very high quality with a volume bias of only about 1%. As such, one may expect the contribution from the EnsPost to be modest, coming mostly from improved reliability by adding spread to the streamflow ensembles.

Figure 6 shows the mean error for the ensemble means and the CRPSS for the postprocessed flow ensembles and raw flow ensembles in reference to the climatology-based flow ensembles. The GFS-based flow ensembles exhibit a conditional bias consistent with the conditional bias of the precipitation ensembles: overforecasting of small events and underforecasting of large events. However, owing to hydrologic persistence or "basin memory," the quality of the flow ensembles declines more slowly than that of the precipitation ensembles. Regarding CRPSS results, the sharp increase in skill between the first and second forecast days is due to the fact that, for the first lead day, the climatology-based flow ensembles too have good skill owing to persistence, which results in reduced skill score for the GFS-based flow ensembles. The comparison of the CRPSS values for the raw flow ensembles and the postprocessed flow ensembles shows that most of the flow forecast skill comes from the MEFP component, with limited impact of EnsPost. The additional improvement by EnsPost is marginal because of small hydrologic biases and uncertainties in this basin. It decreases very fast within the first few days as a reflection of the fast-decaying memory in the initial conditions, noting that the prior observation is a predictor in the EnsPost.

However, as pointed out by Brown (2013), the overall skill of GFS-based postprocessed flow ensembles in reference to climatology-based flows, as well as the relative contributions of the MEFP (with GFS forecasts) and EnsPost components, depend on the basin location (as illustrated in Fig. 7 with basins located in four different RFCs), flow amount,

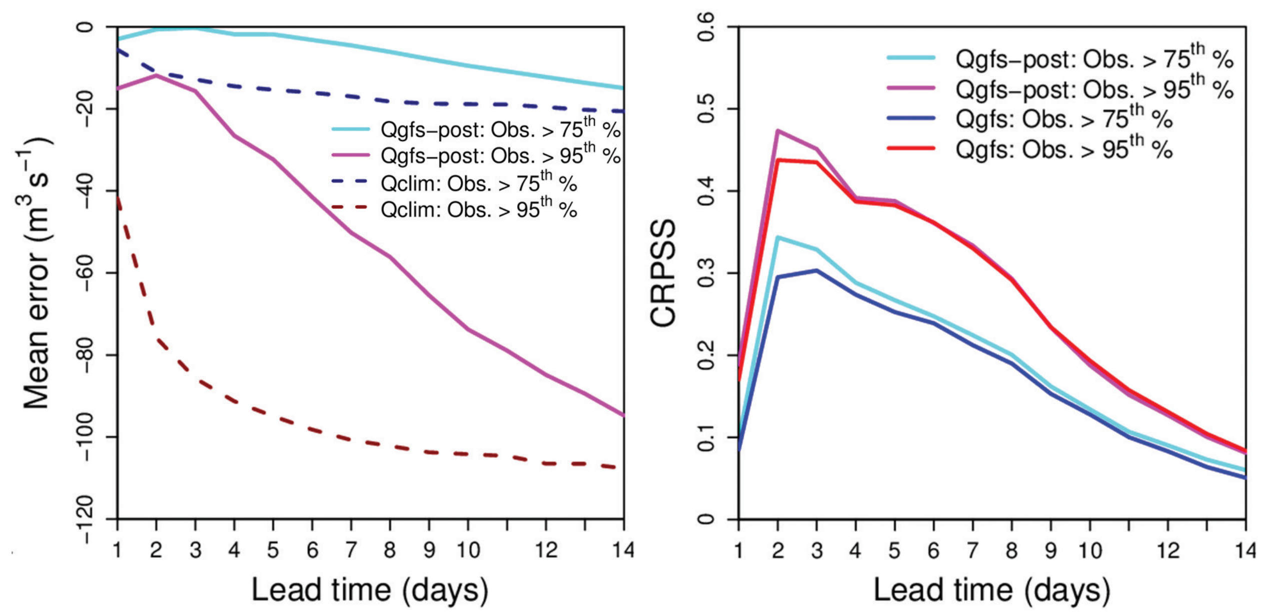

Fig. 6. (left) Mean error of the ensemble means for GFS-based postprocessed flow ensembles ("Qgfs-post") and climatology-based flow ensembles ("Qclim"); (right) CRPSS for the GFS-based postprocessed flow ensembles (Qgfs-post) and GFS-based raw flow ensembles (Qgfs) in reference to the climatology-based flow ensembles (Qclim). The results from dependent validation for the 1979–2005 period for the NFDC1 basin are shown for subsets of forecast-observed pairs relative to the 0.75 and 0.95 nonexceedance probability thresholds (29.6 $m^3 s^{-1}$ and 80.7 $m^3 s^{-1}$, respectively).

and season. In Fig. 8, examples of ensemble traces for two different basins illustrate how MEFP and EnsPost may help to predict a large flow event 5 days in advance but with a peak timing error (top plots), and may produce ensembles with a much reduced spread compared to climatology-based flows but with a low bias tendency (bottom plots).

The performance of EnsPost depends largely on the availability of long-term observed and model-simulated flows and the assumption that the streamflow climatology is stationary over a multidecadal period. If additional stratification of the observed–simulated flow dataset is necessary for parameter estimation to improve the model fit for specific conditions (e.g., snowmelt), the EnsPost will require an even larger dataset for its calibration. In the future, for areas where observed and simulated flow data are available at subdaily scales (6 hourly or hourly), direct modeling of the subdaily flow will be necessary for improved performance. The use of multiple temporal scales of aggregation to improve bias correction at longer ranges is under investigation. Evaluation of other bias-correction techniques (including those used for atmospheric forcings) is also ongoing (e.g., van Andel et al. 2012) to find the best approaches for different forecasting situations and forecast attributes.

Moreover, EnsPost needs to be currently applied without any manual modifications of model states and parameters to maintain the consistency between the real-time ensemble flows and the simulated flows used for its calibration, as well as the EnsPost-generated streamflow hindcasts and verification results. Therefore, for real-time ensemble prediction, the set of model states used in HEFS are generated with a simulation time window long enough to minimize the impact of any modifications previously applied in single-valued forecasting. Obviously, EnsPost needs to evolve along with the data assimilator component to utilize automated DA procedures. Meanwhile, given that the current manual modifications address significant limitations in the operational models and datasets, we recommend analyzing the potential impact of these modifications on the performance of HEFS flow ensembles. Such comprehensive evaluation could offer objective guidance on best operational practices for applying manual modifications and cost-effective transitioning of experimental automated DA capabilities into operational ensemble forecasting.

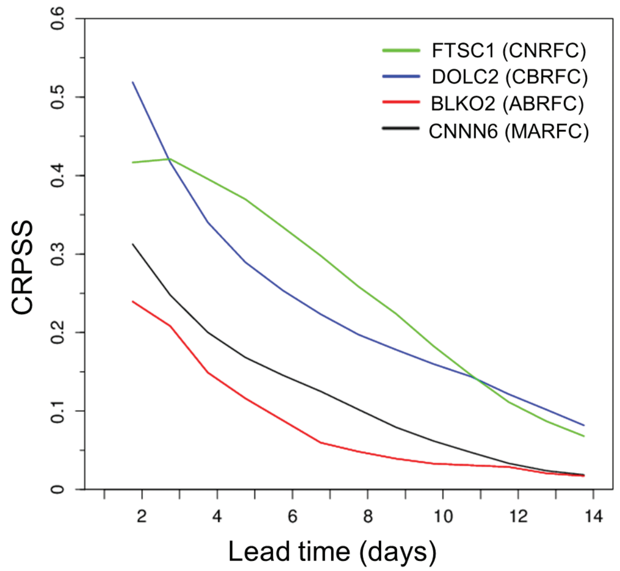

Fig. 7. CRPSS for the GFS-based postprocessed flow ensembles (Qgfs-post) in reference to the climatology-based flow ensembles. The results from dependent validation for the 1979–99 period are shown for all forecast-observed pairs for the following basins: the Eel River at Fort Seward (FTSC1) in California Nevada RFC (CNRFC), the Dolores River (DOLC2) in Colorado Basin RFC (CBRFC), the Chikaskia River near Blackwell (BLKO2) in Arkansas Basin RFC (ABRFC), and the inflow to Cannonsville Reservoir in Middle-Atlantic RFC (MARFC) (see Brown 2013 for details).

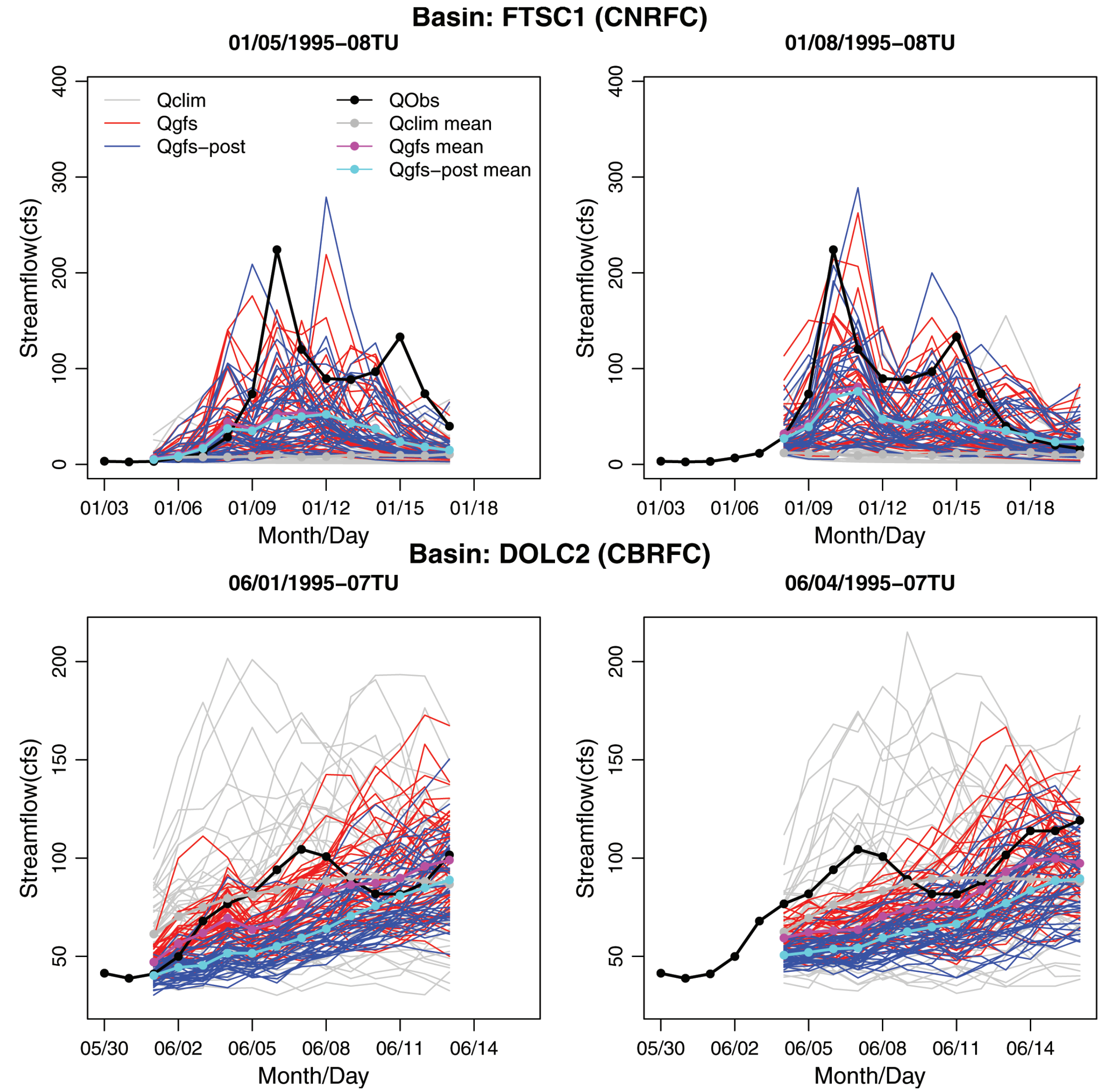

Fig. 8. Daily ensemble traces and ensemble means for the GFS-based postprocessed flow (Qgfs-post), GFS-based raw flow (Qgfs), and climatology-based flow (Qclim) for two different events in (top) the Eel River at Fort Seward (FTSC1) in CNRFC and (bottom) the Dolores River (DOLC2) in CBRFC.

ENSEMBLE VERIFICATION. To evaluate the performance of HEFS for both research and operational forecasting purposes, ensemble verification is required. Key attributes of forecast quality include the degree of bias of the forecast probabilities, whether unconditionally or conditionally upon the forecasts (reliability or Type-I conditional bias) or observations (Type-II conditional bias), the ability to discriminate between different observed events (i.e., to issue distinct probability statements), and skill relative to a baseline forecasting system (Jolliffe and Stephenson 2003; Wilks 2006). Ensemble forecasting systems, such as HEFS, are intended for a wide range of practical applications, such as flood forecasting, river navigation, and water supply forecasting. Therefore, forecast quality needs to be evaluated for a range of observed and forecast conditions in terms of forecast horizon, space–time scale, seasonality, and magnitude of event. The EVS, built on the Ensemble Verification System (Brown et al. 2010; freely available from www.nws.noaa.gov/oh/evs.html), was designed to support conditional verification of forcing and hydrologic ensembles, generated by HEFS, as well as external ensemble forecasting systems. EVS is a flexible, modular, and open-source software tool programmed in Java to allow cost-effective collaborative research and development with academic and private institutions and rapid research-to-operations transition of scientific advances.

Key features of EVS include the following (see Brown et al. 2010 for details):

- the ability to evaluate forecast quality for any continuous numerical variable (e.g., precipitation, temperature, streamflow, river stage) at specific forecast locations (points or areas) and for any temporal scale or forecast lead time;
- the ability to evaluate the quality of an ensemble forecasting system conditional upon many factors, such as forecast lead time, seasonality, temporal aggregation, magnitude of event (defined in various ways, such as exceedance of a real-valued threshold or climatological probability), and values of auxiliary variables (e.g., quality of flow ensembles conditional upon the amount of observed precipitation);
- the ability to evaluate key attributes of forecast quality, such as reliability, discrimination, and

skill, at varying levels of detail, ranging from highly summarized (e.g., skill scores such as CRPSS) to highly detailed (e.g., box plots of conditional errors);
- the ability to aggregate the forecasts in time (e.g., hourly to daily) and to evaluate aggregate performance over a range of forecast locations, either by pooling pairs or computing a weighted average of the verification metrics from several locations;
- generating graphical and numerical outputs in a range of file formats (R scripts are also provided for further analysis and generation of custom graphics);
- the ability to implement a verification study via the graphical user interface (GUI) or to batch process a large number of forecast locations on the command line, using a project file in an XML format (the EVS can also be run within CHPS—e.g., to produce diagnostic verification results for one or multiple hindcast scenarios); and
- the ability to estimate the sampling uncertainty in the verification metrics using the stationary block bootstrap—synthetic realizations of the original paired data are repetitively generated and the verification metrics are computed for each sample to estimate a bootstrap distribution of the verification metrics, from which the percentile confidence intervals are then derived.

EVS is regularly enhanced to address needs from modelers and forecasters as HEFS is being implemented and evaluated across all RFCs and since the Ensemble Verification System is being used in other projects such as HEPEX.

GRAPHICS GENERATOR. Communicating uncertainty information to a wide range of end users represents a challenge. As hydrologic ensemble forecasting is relatively new, much research is needed to define the most effective methods of presenting such information and design decision support system that maximize their utility (Cloke and Pappenberger 2009). Challenges in communicating hydrologic ensembles include how to understand the ensemble forecast information (e.g., value of the ensemble mean, relation between spread and skill), how to use such information (e.g., in coordination with deterministic forecasts), and how to communicate it (e.g., spaghetti plots versus plume charts), even to nonexperts (Demeritt et al. 2010). A variety of practical approaches and products have been presented by Bruen et al. (2010) for seven European ensemble forecasting platforms and by Ramos et al. (2007) and Demeritt et al. (2013) for the European Flood Alert System. Pappenberger et al. (2013) formulated recommendations for effective visualization and communication of probabilistic flood forecasts among experts, acknowledging that there is no overarching agreement and one-size-fits-all solution.

In HEFS, the Graphics Generator (GraphGen), a generic software tool for CHPS, enables forecasters to generate and visualize information for internal decision support during operations as well as disseminate the final products to end users. This tool is expected to be accessed externally through a web service interface, which will allow the uncertainty-quantified forecast and verification information to be tailored to the needs of specific external users. GraphGen includes the functionality of the NWSRFS Ensemble Streamflow Prediction Analysis and Display Program (National Weather Service 2012), such as generation of spaghetti plots, expected value chart to describe the ensemble distribution (minimum, maximum, mean, and standard deviation), exceedance probability bar graph for a few probability categories and for a given product (e.g., monthly volume), and exceedance probability distribution plot using current initial conditions compared to historical simulations (see examples in McEnery et al. 2005). HEFS also needs to provide uncertainty-quantified forecast information on maps for multiple forecast locations (e.g., 90% chance to exceed flood impact thresholds) as well as individual locations (e.g., expected value charts for all forecast lead times). New products to visualize the ensemble information are being evaluated, such as box-and-whisker plots with quantiles from the ensemble distribution, ensemble consistency tables, and visualization of peak timing uncertainty and magnitude uncertainty. In addition, information is needed to link and integrate traditional single-valued forecasts in the context of the estimated uncertainty. Several RFCs make prototype ensemble products and information available to their customers (see www.cnrfc.noaa.gov/index.php?type=ensemble) as well as interfaces that allow custom product generation based on user specifications; this will help to enhance the operational national web interface for AHPS (http://water.weather.gov/ahps/).

Furthermore, verification information needs to be provided along with forecast information to support decision making (Demargne et al. 2009). Similar approaches have been reported by Bartholmes et al. (2009), Renner et al. (2009), and van Andel et al. (2008), where verification results on past performance are provided to forecasters and decision makers. One such example is the web interface for the CBRFC water supply forecasts available from www.cbrfc.noaa.gov/. It enables users to generate customizable datasets and products for both probabilistic forecasts and verification statistics. Such visualization tools provide insights into the strengths and weaknesses of the forecasts and can help users assess potential forecast errors in the real-time forecasts. Along with the probabilistic forecast information, the envisioned HEFS products will include context information (e.g., historical lowest and highest, specific years of interest—El Niño, La Niña, or neutral), recent forecasts and corresponding observations, forecasts from alternative scenarios, as well as historic analogs to real-time forecasts. The selection of analogs (i.e., past forecasts that are analogous to the current forecast) and the display of diagnostic verification statistics from similar conditions provide important contextual information tailored to the specific real-time forecasting situation.

Through customer evaluations of the AHPS website and the NWS Hydrology Program, the NWS has recognized the need to better communicate hydrologic forecast uncertainty information for the end users to understand better and use such information more effectively in their decision making. The NWS, USACE, and Bureau of Reclamation conducted a comprehensive use and needs assessment of the water management community, stressing in particular the need of more detailed information on product skill and uncertainty, guidance for synthesizing the large amount of hydrometeorological information, and training on probabilistic forecasting principles and risk-based decision making (Raff et al. 2013). Increased collaborations between forecasters, scientists (including social and behavioral scientists), and decision makers should help to understand decision processes with uncertainty-based forecasts, develop innovative training and education activities to promote a common understanding, and, ultimately, increase the effectiveness of probabilistic forecasts (Ramos et al. 2013; Demeritt et al. 2013; Pappenberger et al. 2013). To this end, the NWS Hydrology Program and the RFCs are involved in a number of outreach and training activities, as well as ongoing collaboration with the New York City Department of Environmental Protection. Finally, as CHPS and HEFS are based on the Delft-FEWS platform, complementary visualization techniques and decision support systems are expected to be shared within the Delft-FEWS community as well as the broader community to maximize the utility of hydrologic and water resources forecast products and services.

CONCLUSIONS AND FUTURE CHALLENGES. The end-to-end HEFS provides, for short to long range, uncertainty-quantified forecast and verification products that are generated by 1) the MEFP, which ingests weather and climate forecasts from multiple Numerical Weather Prediction models to produce seamless and bias-corrected precipitation and temperature ensembles at the hydrologic basin scales; 2) the Hydrologic Processor, which inputs the forcing ensembles into a suite of hydrologic, hydraulic, and reservoir models; 3) the EnsPost, which models the collective hydrologic uncertainty and corrects for biases in the streamflow ensembles; 4) the EVS, which verifies the forcing and streamflow ensembles to help identify the main sources of skill and bias in the forecasts; and 5) the Graphics Generator, which enables forecasters to derive and visualize products and information from the ensembles. Evaluation of the HEFS through multiyear hindcasting and large-sample verification is currently underway and results obtained so far show positive skill and reduced bias in the short to medium term when compared to climatology-based ensembles and single-valued forecasts. However, the performance varies significantly with, for example, forecast horizons, basin locations, seasons, and magnitudes, which underlines the need for a systematic and comprehensive evaluation of HEFS ensembles across the different RFCs.

Increased skill in forcing forecasts generally translates into increased skill in ensemble streamflow forecasts. As such, the HEFS should utilize the most skillful forcing forecasts at all ranges of lead time. To translate this skill to ensemble streamflow forecasts to the maximum extent, hydrologic uncertainty must be reduced as much as possible. For example, assimilation of all available measurements of streamflow, soil moisture, snow depth, and others would reduce the initial condition uncertainty. Although not implemented in the first version of the HEFS, a number of DA techniques have been developed and/or tested for the Sacramento rainfall-runoff model, the snow accumulation and ablation model, and hydrologic routing models to assimilate real-time observations and adjust model states within the assimilation window (Seo et al. 2003, 2009; Lee et al. 2011; Liu et al. 2012; He et al. 2012). In addition to the data assimilator component, enhancements are also planned to account for the parametric and structural uncertainties in the hydrologic models. As shown by Georgakakos et al. (2004) and Velázquez et al. (2011), combining predictions from different models outperforms individual model predictions as long as model-specific biases can be corrected.

Obviously, the different uncertainty modeling approaches available in the HEFS and in other research and operational systems will need to be rigorously compared via ensemble verification to define optimized systems for operational hydrologic ensemble predictions. Close collaborations between scientists, forecasters, and end users from the atmospheric and hydrologic communities, through projects such as the HEPEX, help support such intercomparison, as well as address the following ensemble challenges:

- seamlessly combine probabilistic forecasts from short to long ranges and from multiple models while maintaining consistent spatial and temporal relationships across different scales and variables;
- include forecaster guidance on forcing input forecasts and hydrologic model operations, especially in the short term;
- improve accuracy of both meteorological and hydrologic models and reduce the cone of uncertainty for effective decision support;
- improve the uncertainty modeling of rare events (e.g., record flooding or drought) when availability of analogous historical events is very limited;
- integrate and leverage conditional uncertainty associated with NWP and human adjusted forecasts of atmospheric forcings;
- improve computing power, database, and data storage, with forecasts becoming available at higher resolution and from an increasing number of models, to produce long hindcast datasets for all forcing inputs and hydrologic outputs for research and operation purposes;
- improve the understanding of how uncertainty and verification information is interpreted and used in practice by different groups (including forecasters and end users) to provide this information in a form and context that is easily understandable and useful to customers; and
- develop innovative training and education activities to fully embrace and practice the ensemble paradigm in hydrology and water resources services and increase the effectiveness of probabilistic forecasts in risk-based decision making.

ACKNOWLEDGMENTS. This work has been supported by the National Oceanic and Atmospheric Administration (NOAA) through the Advanced Hydrologic Prediction Service (AHPS) Program and the Climate Predictions Program for the Americas of the Climate Program Office. The research and development of HEFS over the last decade has involved multiple scientists and forecasters from the Office of Hydrologic Development and the RFCs. The authors would also like to thank Dr. Yuqiong Liu for her valuable contribution and three anonymous reviewers.

REFERENCES

Addor, N., S. Jaun, F. Fundel, and M. Zappa, 2011: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.,* **15,** 2327–2347.

Anderson, E. A., 1973: National Weather Service River Forecast System—Snow accumulation and ablation model. NOAA Tech. Memo. NWS HYDRO-17, 217 pp.

Bartholmes, J. C., J. Thielen, M.-H. Ramos, and S. Gentilini, 2009: The European Flood Alert System EFAS—Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.,* **13,** 141–153.

Bogner, K., and M. Kalas, 2008: Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmos. Sci. Lett.,* **9,** 95–102.

——, and F. Pappenberger, 2011: Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resour. Res.,* **47,** W07524, doi:10.1029/2010WR009137.

Brown, J. D., 2013: Verification of temperature, precipitation and streamflow forecasts from the NWS Hydrologic Ensemble Forecast Service (HEFS): Medium-range forecasts with forcing inputs from the frozen version of NCEP's Global Forecast System. HSL Tech. Rep. to the NWS, 133 pp. [Available online at www.nws.noaa.gov/oh/hrl/hsmb/docs/hep/publications_presentations/Contract_2012-04-HEFS_Deliverable_02_Phase_I_report_FINAL.pdf.]

——, and D.-J. Seo, 2010: A nonparametric postprocessor for bias-correction of hydrometeorological and hydrologic ensemble forecasts. *J. Hydrometeor.,* **11,** 642–665.

——, and ——, 2013: Evaluation of a nonparametric post-processor for bias-correction and uncertainty estimation of hydrologic predictions. *Hydrol. Processes,* **27,** 83–105, doi:10.1002/hyp.9263.

——, J. Demargne, D.-J. Seo, and Y. Liu, 2010: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environ. Modell. Software,* **25,** 854–872.

——, D.-J. Seo, and J. Du, 2012: Verification of precipitation forecasts from NCEP's Short Range Ensemble Forecast (SREF) system with reference to ensemble streamflow prediction using lumped hydrologic models. *J. Hydrometeor.,* **13,** 808–836.

Bruen, M., P. Krahe, M. Zappa, J. Olsson, B. Vehvilainen, K. Kok, and K. Daamen, 2010: Visualizing flood forecasting uncertainty: Some current European EPS platforms—COST731 working group 3. *Atmos. Sci. Lett.,* **11,** 92–99.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP Global ensemble prediction systems. *Mon. Wea. Rev.,* **133,** 1076–1097.

Burnash, R. J. C., 1995: The NWS river forecast system—Catchment modeling. *Computer Models of Watershed Hydrology,* V. P. Singh, Ed., Water Resources Publications, 311–366.

Charba, J. P., D. W. Reynolds, B. E. McDonald, and G. M. 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.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.,* **5,** 243–262.

Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.,* **375,** 613–626.

Coccia, G., and E. Todini, 2011: Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrol. Earth Syst. Sci.,* **15,** 3253–3274, doi:10.5194/hess-15-3253-2011.

Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasting,* **27,** 396–410.

Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Resour. Plann. Manage.,* **111,** 157–170.

Demargne, J., L. Wu, D.-J. Seo, and J. C. Schaake, 2007: Experimental hydrometeorological and hydrologic ensemble forecasts and their verification in the U.S. National Weather Service. Quantification and reduction of predictive uncertainty for sustainable water resources management, IAHS Publ. 313, 177–187.

——, M. Mullusky, K. Werner, T. Adams, S. Lindsey, N. Schwein, W. Marosi, and E. Welles, 2009: Application of forecast verification science to operational river forecasting in the U.S. National Weather Service. *Bull. Amer. Meteor. Soc.,* **90,** 779–784.

——, J. D. Brown, Y. Liu, D.-J. Seo, L. Wu, Z. Toth, and Y. Zhu, 2010: Diagnostic verification of hydrometeorological and hydrologic ensembles. *Atmos. Sci. Lett.,* **11,** 114–122.

Demeritt, D., S. Nobert, H. Cloke, and F. Pappenberger, 2010: Challenges in communicating and using ensembles in operational flood forecasting. *Meteor. Appl.,* **17,** 209–222.

——, ——, ——, and ——, 2013: The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. *Hydrol. Processes,* **27,** 147–157, doi:10.1002/hyp.9419.

Du, J., G. Dimego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, J. Wang, and H. Juang, 2009: Recent upgrade of NCEP short-range ensemble forecast (SREF) system. *Proc. 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting,* Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_153264.htm.]

Fundel, F., S. Jörg-Hess, and M. Zappa, 2012: Long-range hydrometeorological ensemble predictions of drought parameters. *Hydrol. Earth Syst. Sci. Discuss.,* **9,** 6857–6887.

Georgakakos, K., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *J. Hydrol.,* **298** (1–4), 222–241.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble Model Output Statistics and minimum CRPS estimation. *Mon. Wea. Rev.,* **133,** 1098–1118.

Gupta, H. V., K. J. Beven, and T. Wagener, 2005: Model calibration and uncertainty estimation. *The Encyclopedia of Hydrological Sciences,* M. Anderson, Ed., John Wiley and Sons, 2015–2032.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts Mon. Wea. Rev., **132,** 1434–1447.

——, ——, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.,* **87,** 33–46.

——, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.,* **136,** 2620–2632.

——, G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.,* **94,** 1553–1565.

He, M., T. S. Hogue, S. A. Margulis, and K. J. Franz, 2012: An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. *Hydrol. Earth Syst. Sci.,* **16,** 815–831.

He, Y., F. Wetterhall, H. L. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor, 2009: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteor. Appl.,* **16,** 91–101.

——, and Coauthors, 2010: Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: A case study. *Atmos. Sci. Lett.,* **11,** 132–138.

Herr, H. D., and R. Krzysztofowicz, 2005: Generic probability distribution of rainfall in space: The bivariate model. *J. Hydrol.,* **306,** 234–263.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Hopson, T., and P. Webster, 2010: A 1–10 day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003–07. *J. Hydrometeor.,* **11,** 618–641.

Jaun, S., and B. Ahrens, 2009: Evaluation of a probabilistic hydrometeorological forecast system. *Hydrol. Earth Syst. Sci.,* **13,** 1031–1043.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley and Sons, 240 pp.

Kang, T.-H., Y.-O. Kim, and I.-P. Hong, 2010: Comparison of pre- and post-processors for ensemble streamflow prediction. *Atmos. Sci. Lett.,* **11,** 153–159.

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

Lee, H., D.-J. Seo, and V. Koren, 2011: Assimilation of streamflow and in-situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Adv. Water Resour.,* **34,** 1597–1615.

Liu, Y., J. D. Brown, J. Demargne, and D.-J. Seo, 2011: A wavelet-based approach to assessing timing errors in hydrologic predictions. *J. Hydrol.,* **397** (3–4), 210–224.

——, and Coauthors, 2012: Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.,* **16,** 3863–3887.

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.

Möller, A., A. Lenkoski, and T. L. Thorarinsdottir, 2013: Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quart. J. Roy. Meteor. Soc.,* **139,** 982–991, doi:10.1002/qj.2009.

Montanari, A., and G. Grossi, 2008: Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resour. Res.,* **44,** W00B08, doi:10.1029/2008WR006897.

National Research Council of the National Academies, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts.* The National Academies Press, 124 pp.

National Weather Service, cited 2012: National Weather Service River Forecast System (NWSRFS) user manual documentation. National Weather Service. [Available online at www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/htm/xrfsdocpdf.php.]

Pappenberger, F., J. Bartholomes, J. Thielen, H. Cloke, R. Buizza, and A. de Roo, 2008: New dimensions in early flood warming across the globe using grand-ensemble weather predictions. *Geophys. Res. Lett.,* **35,** L10404, doi:10.1029/2008GL033837.

——, E. Stephens, J. Thielen, P. Salamon, D. Demeritt, S. J. van Andel, F. Wetterhall, and L. Alfieri, 2013: Visualizing probabilistic flood forecast information: Expert preferences and perceptions of best practice in uncertainty communication. *Hydrol. Processes,* **27,** 132–146, doi:10.1009/hyp.9253.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. ECMWF Tech. Memo. 548, 47 pp.

Philpott, A. W., P. Wnek, and J. D. Brown, cited 2012: Verification of ensemble river forecasts at the Middle Atlantic River Forecast Center. [Available online at https://ams.confex.com/ams/92Annual/webprogram/Paper199532.html.]

Pyke, G., and J. Porter, cited 2012 Forecast-based operations support tool for the New York City water supply system. [Available online at http://fallmeeting.agu.org/2012/files/2012/12/2012-AGU-Fall-Meeting-Poster-Pyke-11x17.pdf.]

Raff, D., L. Brekke, K. Werner, A. Wood, and K. White, 2013: Short-term water management decisions: User needs for improved climate, weather, and hydrologic information. U.S. Army Corps of Engineers, Bureau of Reclamation, and National Oceanic and Atmospheric Administration Tech. Rep. CWTS 2013-1, 256 pp. [Available online at www.ccawwg.us/docs/Short-Term_Water_Management_Decisions_Final_3_Jan_2013.pdf.]

Ramos, M.-H., J. Bartholmes, and J. Thielen, 2007: Development of decision support products based on ensemble weather forecasts in the European flood alert system. *Atmos. Sci. Lett.,* **8,** 113–119.

——, S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.* **17,** 2219–2232, doi:10.5194/hess-17-2219-2013.

Reggiani, P., M. Renner, A. H. Weerts, and P. A. H. J. M. van Gelder, 2009: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resour. Res.,* **45,** W02428, doi:10.1029/2007WR006758.

Regonda, S., D.-J. Seo, B. Lawrence, J. D. Brown, and J. Demargne, 2013: Short-term ensemble streamflow forecasting using operationally-produced single-valued streamflow forecasts—A Hydrologic Model Output Statistics (HMOS) approach. *J. Hydrol.,* **497,** 80–96, doi:10.1016/j.jhydrol.2013.05.028.

Renner, M., M. G. F. Werner, S. Rademacher, and E. Sprokkereef, 2009: Verification of ensemble flow forecasts for the River Rhine. *J. Hydrol.,* **376** (3–4), 463–475.

Saha, S., and Coauthors, 2013: The NCEP Climate Forecast System Version 2. *J. Climate,* in press.

Schaake, J., and Coauthors, 2007a: Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol. Earth Syst. Sci. Discuss.,* **4,** 655–717.

——, T. M. Hamill, and R. Buizza, 2007b: HEPEX: The Hydrological Ensemble Prediction Experiment. *Bull. Amer. Meteor. Soc.,* **88,** 1541–1547.

——, and Coauthors, 2010: Summary of recommendations of the first workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Météo-France, Toulouse, France, 15–18 June 2009. *Atmos. Sci. Lett.,* **11,** 59–63.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.,* in press.

Schellekens, J., A. H. Weerts, R. J. Moore, C. E. Pierce, and S. Hildon, 2011: The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales. *Adv. Geosci.,* **29,** 77–84.

Seo, D.-J., V. Koren, and N. Cajina, 2003: Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *J. Hydrometeor.,* **4,** 627–641.

——, H. Herr, and J. Schaake, 2006: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol. Earth Syst. Sci. Discuss.,* **3,** 1987–2035.

——, L. Cajina, R. Corby, and T. Howieson, 2009: Automatic state updating for operational streamflow forecasting via variational data assimilation. *J. Hydrol.,* **367,** 255–275.

Singla, S., J.-P. Céron, E. Martin, F. Regimbeau, M. Déqué, F. Habets, and J.-P. Vidal, 2012: Predictability of soil moisture and river flows over France for the spring season. *Hydrol. Earth Syst. Sci.,* **16,** 201–216.

Smith, P. J., K. Beven, A. Weerts, and D. Leedal, 2012: Adaptive correction of deterministic models to produce accurate probabilistic forecasts. *Hydrol. Earth Syst. Sci.,* **16,** 2783–2799.

Solomatine, D. P., and D. L. Shrestha, 2009: A novel method to estimate model uncertainty using machine learning techniques. *Water Resour. Res.,* **45,** W00B11, doi:10.1029/2008WR006839.

Thielen, J., J. Schaake, R. Hartman, and R. Buizza, 2008: Aims, challenges and progress of the Hydrological Ensemble Prediction Experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. *Atmos. Sci. Lett.,* **9,** 29–35.

Thirel, G., F. Regimbeau, E. Martin, J. Noilhan, and F. Habets, 2010: Short- and medium-range hydrological ensemble forecasts over France. *Atmos. Sci. Lett.,* **11,** 72–77.

Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manage.,* **6,** 123–137.

van Andel, S. J., R. K. Price, A. H. Lobbrecht, F. van Kruiningen, and R. Mureau, 2008: Ensemble precipitation and water-level forecasts for anticipatory water-system control. *J. Hydrometeor.,* **9,** 776–788.

——, A. Weerts, J. Schaake, and K. Bogner, 2012: Post-processing hydrological ensemble predictions intercomparison experiment. *Hydrol. Processes,* **27,** 158–161, doi:10.1002/hyp.9595.

Van den Bergh, J., and E. Roulin, 2010: Hydrological ensemble prediction and verification for the Meuse and Scheldt basins. *Atmos. Sci. Lett.,* **11,** 64–71.

Velázquez, A., T. Petit, A. Lavoie, M.-A. Boucher, R. Turcotte, V. Fortin, and F. Anctil, 2009: An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting. *Hydrol. Earth Syst. Sci.,* **13,** 2221–2231.

——, F. Anctil, M.-H. Ramos, and C. Perrin, 2011: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Adv. Geosci.,* **29,** 33–42.

Weerts, A. H., H. C. Winsemius, and J. S. Verkade, 2011: Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrol. Earth Syst. Sci.,* **15,** 255–265.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP Global Operational Forecast System. *Tellus,* **60A,** 62–79.

Welles, E., S. Sorooshian, G. Carter, and B. Olsen, 2007: Hydrologic verification: A call for action and collaboration. *Bull. Amer. Meteor. Soc.,* **88,** 503–511.

Werner, K., D. Brandon, M. Clark, and S. Gangopadhyay, 2005: Incorporating medium-range numerical weather model output into the ensemble streamflow prediction system of the National Weather Service. *J. Hydrometeor.,* **6,** 101–114.

Werner, M., M. van Dijk, and J. Schellekens, 2004: DELFT-FEWS: An open shell flood forecasting system. *Proceedings of the 6th International Conference on Hydroinformatics,* S.-Y. Liong, K.-K. Phoon, and V. Babović, Eds., World Scientific, 1205–1212.

——, M. Cranston, T. Harrison, D. Whitfield, and J. Schellekens, 2009: Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteor. Appl.,* **16,** 13–22.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 648 pp.

——, and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.,* **135,** 2379–2390.

Wood, A. W., and J. C. Schaake, 2008: Correcting errors in streamflow forecast ensemble mean and spread. *J. Hydrometeor.,* **9,** 132–148.

Wood, E. F., X. Yuan, and J. K. Roundy, 2011: Enhancing hydrological seasonal forecast by downscaling CFSv2. *36th Annual Climate Diagnostics and Prediction Workshop,* Fort Worth, TX, NOAA, 110–115.

Wu, L., D.-J. Seo, J. Demargne, J. Brown, S. Cong, and J. Schaake, 2011: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast via meta-Gaussian distribution models. *J. Hydrol.,* **399** (3–4), 281–298.

Yuan, X., E. Wood, J. Roundy, and M. Pan, 2013: CFSv2-based seasonal hydroclimatic forecasts over conterminous United States. *J. Climate,* **26,** 4828–4847.

Zappa, M., and Coauthors, 2010: Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. *Atmos. Sci. Lett.,* **11,** 83–91.

——, F. Fundel, and S. Jaun, 2012: A 'Peak-Box' approach for supporting interpretation and verification of operational ensemble peak-flow forecasts. *Hydrol. Processes,* **27,** 117–131, doi:10.1002/hyp.9521.

Zhao, L., Q. Duan, J. Schaake, A. Ye, and J. Xia, 2011: A hydrologic post-processor for ensemble streamflow predictions. *Adv. Geosci.,* **29,** 51–59.

Abstract

NOAA's National Weather Service (NWS) is implementing a short- to long-range Hydrologic Ensemble Forecast Service (HEFS). The HEFS addresses the need to quantify uncertainty in hydrologic forecasts for flood risk management, water supply management, streamflow regulation, recreation planning, and ecosystem management, among other applications. The HEFS extends the existing hydrologic ensemble services to include short-range forecasts, incorporate additional weather and climate information, and better quantify the major uncertainties in hydrologic forecasting. It provides, at forecast horizons ranging from 6 h to about a year, ensemble forecasts and verification products that can be tailored to users' needs.

Based on separate modeling of the input and hydrologic uncertainties, the HEFS includes 1) the Meteorological Ensemble Forecast Processor, which ingests weather and climate forecasts from multiple numerical weather prediction models to produce bias-corrected forcing ensembles at the hydrologic basin scales; 2) the Hydrologic Processor, which inputs the forcing ensembles into hydrologic, hydraulic, and reservoir models to generate streamflow ensembles; 3) the hydrologic Ensemble Postprocessor, which aims to account for the total hydrologic uncertainty and correct for systematic biases in streamflow; 4) the Ensemble Verification Service, which verifies the forcing and streamflow ensembles to help identify the main sources of skill and error in the forecasts; and 5) the Graphics Generator, which enables forecasters to create a large array of ensemble and related products. Examples of verification results from multiyear hindcasting illustrate the expected performance and limitations of HEFS. Finally, future scientific and operational challenges to fully embrace and practice the ensemble paradigm in hydrology and water resources services are discussed.