Journal of Hydrology X 11 (2021) 100073

Contents lists available at ScienceDirect

Journal of Hydrology X

journal homepage: www.sciencedirect.com/journal/journal-of-hydrology-x

Techniques to evaluate the modifier process of National Weather Service flood forecasts

Zhipeng Zhu [a], Asphota Wasti [a], Trent Schade [b], Patrick A. Ray [a,*]

[a] College of Engineering and Applied Science, University of Cincinnati, Cincinnati, OH, United States
[b] Ohio River Forecast Center, National Weather Service, Wilmington, OH, United States

ARTICLE INFO

Keywords:
Flood forecast uncertainty
Operational hydrology
ANOVA
Data visualization
Heat maps
National Weather Service

ABSTRACT

The operational hydrologists of the United States' National Weather Service (NWS) develop river forecasts as guidance for those at risk of flood damage and update those flood forecasts in real-time as more information becomes available. To do so they rely on experience and intuition to adjust the inputs, state variables, and parameters of hydrologic models. NWS hydrologists use the term "modifiers" to refer collectively to these adjustments. This paper demonstrates the development and application of tools (statistical and graphical) to aid operational hydrologists in the achievement of accurate flood forecasts. Analysis of variance (ANOVA) identifies the relative contribution to forecast uncertainty of each modifier. Heat map visualizations illustrate for operational hydrologists the basin, lead-time, and season-specific effects of their modifiers choices. The tools provide operational hydrologists with insight into which of three commonly applied modifiers (precipitation, soil moisture, and unit hydrograph shape) are most likely to provide improvement in flood forecast accuracy. The tools are demonstrated for a case study of four watersheds within in the Ohio River Valley, using data for flood events sampled from 1990 to 2018. The findings of this research show that operational hydrologists in the Ohio River Basin would do well apply no modifiers in the winter (leaving hydrologic input variables and parameters at baseline values). And though the forecast might be improved by real-time adjustments to the unit hydrograph in summer months, recommendations for particular unit hydrograph modification levels cannot be made with confidence. These findings call into question the modifier adjustment program as a standard process. In the evaluated cases, modifiers do not systematically improve flood forecasts. Improvement may be more efficiently achieved through better calibration of hydrologic models or techniques for reduction of precipitation uncertainty.

1. Introduction

Floods affect more people globally than any other type of natural disaster, inflicting devastating damages on human life and property (IFRC, 2015). In 2016, floods claimed 26 lives in the Ohio River Valley. Same year at a national scale, flooding caused 126 fatalities and over $10 billion in damages (NOAA, 2016a). The emergency management community, including federal agencies, state organizations and local police, fire, and rescue, coordinate and respond to these natural disasters. The National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) provides forecasts and warnings before floods start. The emergency management community relies on NWS flood forecasts when allocating resources and mobilizing response. Increased accuracy and timeliness of forecasts would allow better targeted, more efficient mobilization of preventative measures and emergency response, which can be used by stakeholders and decision makers to respond to floods before they occur, saving lives and protecting property. Furthermore, investments in improvements to flood forecasts are cost effective and of low socio-environmental impact relative to the construction of new water infrastructure for flood protection (such as dams or levees) or modification of infrastructure operation rules and re-issuing of infrastructure control manuals (Butts et al., 2007).

The NWS provides real-time flood forecasts in collaboration with other agencies: the United States Geological Survey (USGS), US Bureau of Reclamation (USBR), US Army Corps of Engineers (USACE), and the Environmental Protection Agency (USEPA), among others. Thirteen NWS River Forecast Centers (RFCs) are responsible for providing the 5-

* Corresponding author.
E-mail address: patrick.ray@uc.edu (P.A. Ray).

https://doi.org/10.1016/j.hydroa.2020.100073
Received 26 June 2020; Received in revised form 8 December 2020; Accepted 28 December 2020
Available online 31 December 2020

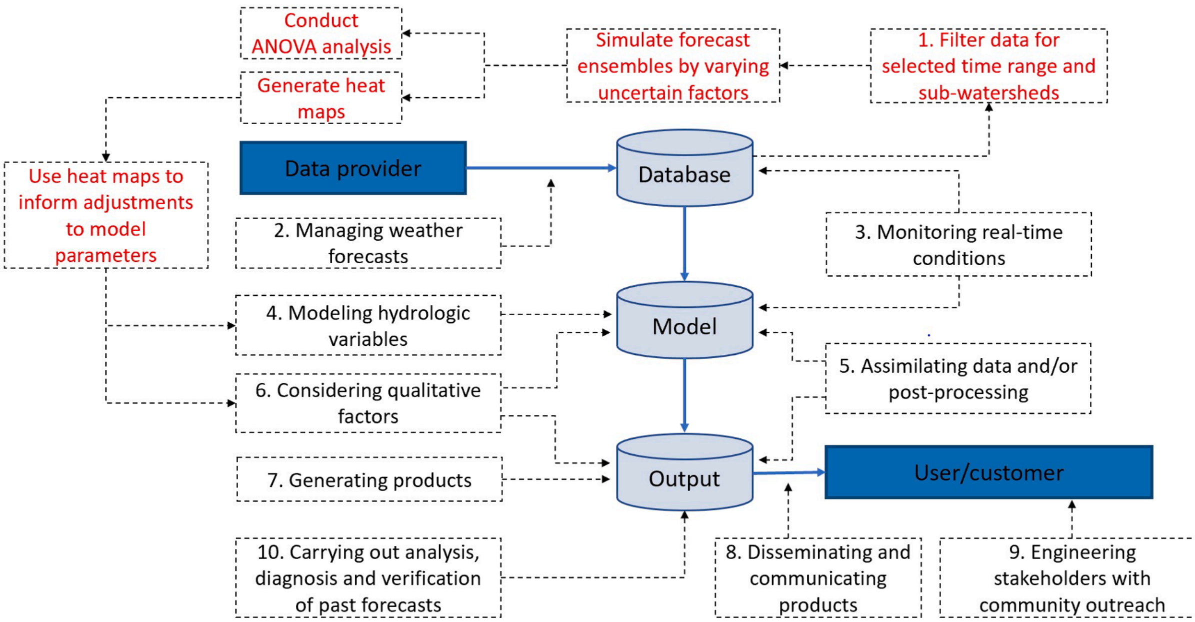

Fig. 1. Operational tasks and experimental design (modification of Pagano et al., 2016).

day streamflow forecasts for rivers throughout the country. The RFCs use the NWS Community Hydrologic Prediction System (CHPS; Adams and Pagano, 2016), which is based on the Flood Early Warning System (FEWS; Deltares, 2018). FEWS is a platform for the construction of operational forecasting systems and includes the ability to flexibly integrate third-party models and data (Werner et al., 2013). The NWS communicates river forecast products including the magnitude and uncertainty of floods or droughts to end users through the NOAA/NWS Advanced Hydrologic Prediction Service (AHPS).

Despite the importance of NWS flood forecasts to national interests, relatively little progress has been made to systematically verify forecasts, and further scientific research is needed to develop standard processes for hydrologic forecast verification (Welles et al., 2007; Brown et al., 2010; Zalenski et al., 2017). Channeling G.P. Box's (1979) aphorism "all models are wrong, some are useful," NWS hydrologists prioritize activities that promote understanding of error sources and adapt model parameters and configurations as those errors arise.

Research regarding NWS streamflow forecasts shows room for improvement in forecasts, especially at longer lead-times (defined as the time between flood forecast start time and the time of peak streamflow) and with conditions of above flood-stage water levels (Welles et al., 2007). Welles & Sorooshian (2009) demonstrated that improved estimates of antecedent hydrologic conditions would be especially useful at shorter lead-times, and improved precipitation forecasts would have the greatest positive effect on flood forecast accuracy at longer lead-times, especially above flood-stage forecasts. Post-processing is another method to reduce forecast uncertainty, yet the literature is scarce. Kang et al. (2010) used two case studies in Korea to demonstrate that the use of post-processing methods can effectively reduce the forecast uncertainty. Roulin and Vannitsem (2015) found that post-processing can largely correct the errors in parameter values. Better basin-specific calibrations of hydrologic and hydraulic routing models would be useful in most cases; however, calibrations targeting accuracy at high flows tend to sacrifice fidelity during periods of low flow, and vice versa. Because the same hydrologic models are used to answer a number of questions regarding flood and drought, calibration target compromises are required.

Operational hydrology standards of practice in the NWS encourage forecasters to apply expert judgement to modify streamflow forecasts in real time, as new data become available. The FEWS software contains a "factors" mechanism to adjust forecast models, inputs, and state variables. Previous evaluations and verifications of NWS short-range flood forecasts (e.g., Liu & Gupta, 2007; Zappa et al., 2011), have not systematically evaluated the impact of human decision (i.e., operational hydrologist expert judgement) on forecast quality. In part to address this concern, the NWS has identified the need for better visualizations of the tendency of operational hydrologist modifications to improve forecast accuracy across a range of representative conditions. The goal of this paper is to evaluate the uncertainty in the NWS's forecast model and develop a visualization tool to help operational hydrologists better understand the modifier process of the NWS.

1.1. *Operational Hydrology*

In general, hydrologic science can be used for long-term planning purposes (such as water resources infrastructure design and management), or short-term operational purposes (such as the development of real-time forecasts). Excellent texts are available to provide the reader with background on hydrology for planning and management (e.g., Viessman and Lewis, 2003; Dingman, 2014), and a representative snapshot of hydrologic models available for water systems planning under uncertainty is presented in Ray and Brown (2015).

Different from hydrology for planning purposes, operational flood forecasting systems are real-time and ever-evolving. The World Meteorological Organization (WMO) and Global Water Partnership (GWP) (2013) explain that, under operational conditions (e.g., during flood warning operations), the role of a forecast service is to collect and process data, run the forecast models, and then provide the forecast products to end users. Under standby conditions, the role of a forecast service is to maintain and improve the performance of the modeling system. The focus of this paper is on enhancements to operational hydrological procedures.

There are four main components of operational forecasting (Zappa et al., 2011): (1) numerical weather predictions (NWP); (2) hydrological initial conditions; (3) flood prediction systems; and (4) warnings for end users. Pagano et al. (2016) investigated 19 forecasting systems and identified three main types. First, passive systems run automatic simulations without human adjustment. Second, in observant systems, humans have little interaction with the model except for considering adjustments in model output when generating public products. Examples of observant systems include the European Flood Awareness System (EFAS), the fledgling global offshoot (GloFAS), and the Flood

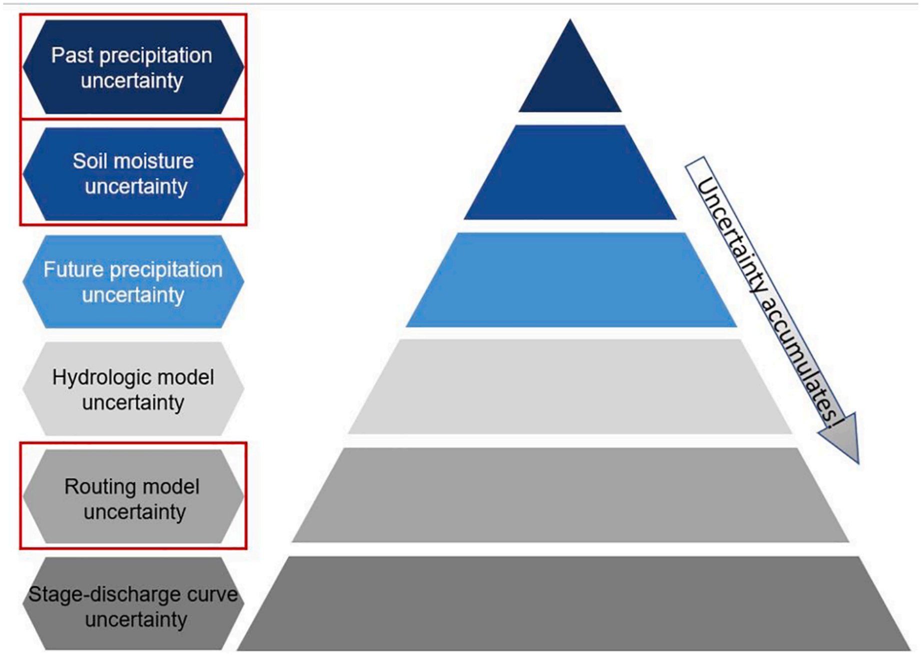

Fig. 2. The cascade of uncertainty in real-time flood forecasting (after Wilby & Dessai, 2010).

Forecasting Center (FFC) in the United Kingdom (UK), which follows the observant paradigm with a higher level of automation. Third, in engaged systems, humans use expertise frequently in real-time forecasting processes to adjust the model. Operational forecasting in the NWS is most nearly of this type. Though humans tend to have a pessimistic view of models–a tendency to think models are worse than they really are (Skitka et al., 1999), expert forecasters are different. They demonstrate the ability to interpret real-world problems using forecast models (Pliske et al., 2004), and they make productive use of the information provided. NWS operational hydrologists actively accommodate model uncertainties. Instead of narrowly interpreting model results as "right" or "wrong", they interpret model results as "likely or not" to provide actionable information to stakeholders.

Fig. 1, modified from Pagano et al. (2016), shows the typical workflow used by operational hydrologists, and the experimental design of this study. The modifications to conventional operational hydrologic processes proposed by this study are presented in red font. This study supplements the conventional operational hydrologic process by: 1) sampling flood events from a database of selected sub-watersheds; 2) generating forecast ensembles using systematically-varied factor levels; and 3) performing ANOVA analysis to decompose forecast uncertainty by factor before developing heat maps to visualize forecast uncertainty under a wide range of input conditions.

1.2. *Uncertainty in Flood Forecasts*

To ensure the usefulness of flood forecasts in disaster prevention and emergency response, accuracy is important, and understanding uncertainty is crucial. Following a process to: 1) define the source of uncertainty; 2) quantify the uncertainty; and 3) evaluate the uncertainty can lead to a reduction of uncertainty (Butts et al, 2007). Todini's (2008) definition of "predictive uncertainty" in flood forecasts as the probability of any future (real) value conditional upon all the knowledge and information achievable through a learning process up to present is adopted here.

Unfortunately, uncertainties in flood forecasts build on each other. While each uncertainty may be relatively manageable on its own, in a cascade of uncertainty, the challenge of producing an accurate flood forecast magnifies. Fig. 2 illustrates this point. Red boxes identify the uncertainties in Fig. 2 that are of primary concern to this study: 1) precipitation; 2) antecedent soil moisture; and 3) the shape of the unit hydrograph. Uncertainty in future precipitation, hydrologic model calibration, and stage-discharge curve relationships are important, but not included among the primary modifiers attended to by NWS operational hydrologists, and therefore are outside of the current scope.

As shown in Fig. 2, a reasonable starting point for flood forecasts is the estimate of the amount of precipitation that has occurred in the past five days, and which is at some stage in the process of impacting streamflow soon to be observed at the station of interest. The precipitation of the past five days is uncertain because each point estimate of precipitation is uncertain (measurement uncertainty), and methods for interpolating between point observations of precipitation depth are imperfect (Kitzmiller et al., 2013). Satellite-based remote-sensed measurements and ground-based radar carry their own uncertainties (Baeck and Smith, 1998; Kitzmiller et al., 2013; Klazura et al., 1999; Smith et al., 1996; Young et al., 1999).

The precipitation that will fall in the coming five days is, of course, not perfectly knowable because: 1) it is derived from imperfect models (e.g., Global Ensemble Prediction System (GEPS; Buizza et al., 2005; Candille et al., 2007) of the Meteorological Services (MSC) of Canada, United States National Center for Atmospheric Research (NCAR; Whitaker et al., 2008), European Centre for Medium-Range Weather Forecasts (Janiskova et al., 2018); and 2) because there is uncertainty in the best way to combine the output of the sampled models (Habib and Qin, 2013). Antecedent soil moisture is estimated using a continuous soil moisture model calibrated over the course of ten years of observation (NOAA, 2010). The hydrologic model used by the NWS, SAC-SMA (NOAA, 2016b), has uncertainty in its energy balance and water balance (Najafi et al., 2011), and the routing model, which is often one-size-fits-all (Viessman and Lewis, 2003), estimates the time to peak flow only approximately. Finally, the stage discharge curve, which translates flow into depth, is subject to imperfect estimates of bathymetry, and often is not derived for the particular basin under examination (Fread, 1973). Details on each of these factors will be provided in this section below.

At the time a forecast is made, errors are contained in: 1) model input, 2) model states or 3) model parameterization. First, uncertain estimates of past precipitation and future precipitation are used, as shown at the top of Fig. 2. Uncertainty in past precipitation uncertainty is caused by measurement error and imperfect interpolation procedures. For estimates of future precipitation, the NWS relies on a network of sensors and gauges along with numerical weather modeling to provide the precipitation inputs that drive the hydrologic model. NWS operational hydrologists are familiar with the error inherent in these inputs (Kitzmiller et al., 2013) and apply expert judgement to adjust for the errors.

The NWS hydrologic model retains a set of state variables through time. These variables can be examined in the model at any point in time to understand the simulated conditions. Most often, it is the soil moisture state variables that are examined. In the NWS continuous soil moisture model, the state changes in time in response to precipitation and temperature inputs at rates determined by a set of parameters. Those parameters are assigned during a model calibration process during which model performance is evaluated against historical data (Gupta et al., 2003). Because NWS operational hydrologists are also involved in the calibration process, as a group they are aware of the errors in model states and capable of applying expert judgement when model states poorly represent the actual conditions of the soil moisture. Of note: because the priority of the NWS is accuracy during high flow events, the soil moisture model calibration tolerates systematic overestimates of (biases in) soil moisture, so that high flow events would not be underestimated.

The hydrologic model adds its own uncertainty because of an inability to satisfactorily represent the complexities of a real-world watershed. NWS operational hydrologists are often able to identify when the simplifying assumptions in the hydrologic model may be violated. For example, the three assumptions inherent in lumped hydrologic modeling are that rainfall occurs: 1) for a duration of one complete time-step; 2) at a steady intensity; and 3) uniformly over the full watershed. When these idealized assumptions are violated, as they nearly always must be, NWS operational hydrologists apply expert judgement to adjust the model. As both precipitation and soil moisture are inputs of the hydrologic model, uncertainty in the value of each accumulates in the hydrologic modeling process.

Uncertainty in the rainfall-runoff from the hydrologic model then flows into the hydraulic routing process. The NWS uses a 6-hour unit hydrograph for every basin regardless of size or slope. Although simple and efficient, the assumptions underlying unit hydrographs are often violated, adding to total forecast uncertainty. In sum, after the stage-discharge transition, the overall uncertainty is embedded in the final flood forecast product.

There are no well-accepted guidelines on quantifying uncertainty within flood forecast systems (Liu and Gupta, 2007) for deterministic hydrologic models. Krzysztofowicz (1999) introduced a Bayesian forecasting system that decomposes forecast uncertainty into input uncertainty and hydrologic uncertainty, which can then be quantified independently and integrated into a Bayesian distribution. Wani et al. (2017) presented a non-parametric method to quantify residual uncertainty, which acts as a post-processor on model forecasts and generates a residual uncertainty distribution. Boelee et al. (2019) identified two methods in general to quantify uncertainty: 1) statistical methods (which calculate past model errors for specific conditions as an estimation of future model uncertainty); and 2) ensemble methods (which create ensembles of forecasts to understand forecast spread). Although in theory either statistical methods or ensemble-based methods could be used to estimate uncertainty in flood forecasts, it is difficult to identify the method most appropriate in any individual case due to the variety and complexity in flood forecasting and warning systems (Boelee et al., 2019). This paper therefore uses both.

This paper discusses past precipitation, soil moisture and flood routing uncertainty, but does not directly deal with uncertainty in future precipitation, hydrologic model calibration, or stage-discharge curves. Ensemble-based methods were used to generate forecast experiments, and (ANOVA) determined the relative sensitivity of the flood forecast to the three considered uncertain factors. Existing literature demonstrates the effectiveness of Analysis of Variance (ANOVA) to decompose aggregate forecast uncertainty into the contributions of model elements (Bosshard et al., 2013; Addor et al., 2014; Antonetti and Zappa, 2018). Heat maps illustrate the response of the generated ensembles to the same three uncertain factors explored using ANOVA techniques, and the resulting graphics provide quick references to understand uncertainty in the generated ensembles. Detailed background on uncertainty in precipitation, soil moisture, unit hydrograph and hydrologic model calibration can be found in Supplemental Information Section 1.

Table 1
General levels of uncertainty in factors for operational hydrologic forecasts.

Precipitation		Soil Moisture		Unit Hydrograph	
1	Least	1	Least	1	Baseline
2	Lesser	2	Less	2	Left shift
3	Less	3	Baseline	3	Right shift
4	Baseline	4	More	4	High pass
5	More	5	Most	5	Low pass
6	Still More			6	Dispersion 1
7	Most			7	Dispersion 2

Table 2
Detailed description of unit hydrograph levels.

Factor level	Description	Interpretation
1	Baseline	Include the original unit hydrograph ordinates
2	Left shift	Drop left-most ordinate
3	Right shift	Add a zero-value ordinate to the left side
4	High pass	Replace lower valued half of ordinates with 0
5	Low pass	Replace upper valued half of ordinates with 0
6	Dispersion 1	Apply a 3-ordinate moving average smoothing
7	Dispersion 2	Apply a 5-ordinate moving average smoothing

2. General Methodology

2.1. Analysis of Variance (ANOVA) for flood forecasts – Experimental design

ANOVA is a statistical procedure in which the total variation in a measured response is partitioned into components, which can be attributed to recognizable sources of variation (Milton and Arnold, 1990). In this study, the sources/factors for variation between observed and simulated river streamflow are the modifiers of primary concern to NWS operational hydrologists: Precipitation (*P*), Soil Moisture (*S*), and Unit Hydrograph (*H*). Table 1 summarizes the three factors (modifiers) and their experimental levels. Each factor has different predefined levels, which represent the conditions for the experiment. The factor *P* is separable into 7 unique levels ($i = 7$) representing the percentage increase/decrease from the baseline value. The factor *S* ranges from completely dry to wet and has 5 unique levels ($j = 5$). Factor *H* represents the shift in ordinates of the Unit Hydrograph: its base line, horizontal shift, flexibility in horizontal directions and flexibility in vertical direction, and are defined by 7 unique levels ($k = 7$). Details of unit hydrograph levels are provided in Table 2. A figure represents different shapes of unit hydrograph is provided in Fig. S1.

The percent error was first computed for each generated ensemble as shown in Eq. (1):

$$E_p = \frac{|(Q_{sim} - Q_{obs})|}{Q_{obs}} * 100\% \tag{1}$$

where, E_p is the percent error for each forecast ensemble, Q_{sim} (Q represents peak flow volume) is the simulated flow, Q_{obs} is the observed flow. In order to better satisfy the normality requirement of the ANOVA analysis, a "box-cox" transformation was performed on E_p. A detailed description of the procedure for a "box-cox" transformation can be found in Box and Cox (1964). Then the ANOVA analysis was performed

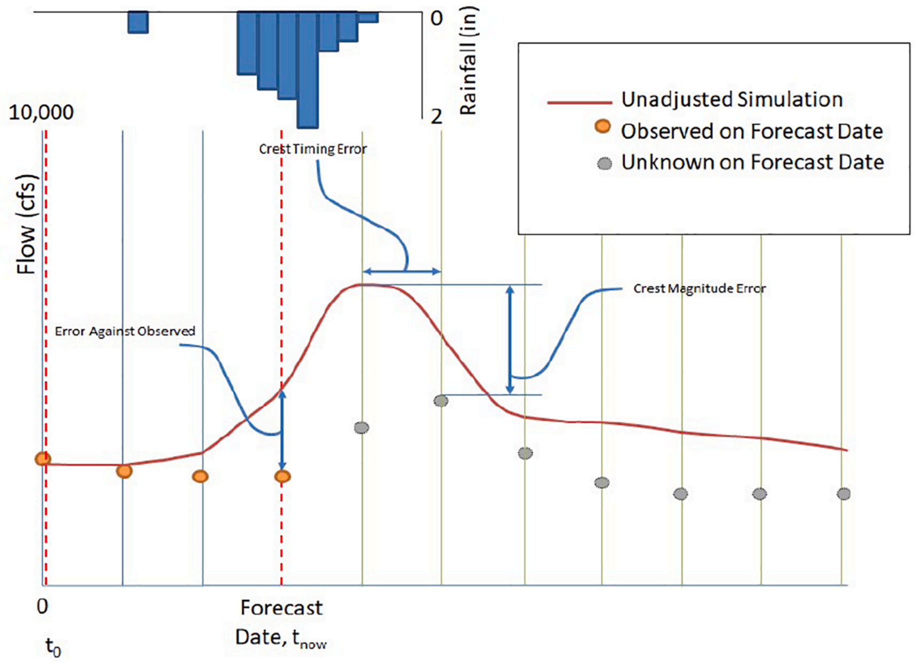

Fig. 3. OHRFC river forecast hydrograph.

on the transformed data of each flood forecast event at different lead-times. Details on the application of standard ANOVA procedures to this study can be found in Supplemental Information Section 3.1.

2.2. Visualizing Uncertainty

Hierarchical displays are promising (e.g., dimensional stacking and pixel display) (Keim, 2002; Žilinskas et al., 2013), in which some features in the plot are embedded in other features to present multi-dimensional data in a 2D basis. Heat maps as visualization aids have existed for over a century. Heat maps are intuitively effective displays of multi-dimensional data, with the *x* and *y* axes representing two dimensions and colored rectangular tiles representing a 3rd dimension corresponding to the values of the data matrix. By combining multiple heat maps in a panel matrix plot, it is possible to show greater dimensions of data in a limited space.

In this paper, the *x* axis of a heat map was used to present the seasonality of flood events (monthly from January to December), and the *y* axis was used to present model parameter levels. Because for each event and lead time the precipitation modifier has 7 different levels that can be adjusted, the soil moisture modifier has 5 levels, and the unit hydrograph modifier has 7 levels, there are 245 possible combinations of adjustments in total. Thus, the simulation results for all 245 ensembles (combinations) were generated using the NWS forecast model. To identify the best forecasts (combinations of parameter and input levels) in each reproduced historical case, only ensembles with the lowest percent error were displayed in heat map for each event. Colored tiles were then used to present the percent distribution of factor levels within each factor type among the best forecasts (see Eq. (1) for the percent error calculation). For instance, in January, the distribution of percent of precipitation factor levels is: 70% of events used precipitation level 4, 20% used level 6 and 10% used level 7. Because only forecasts with the lowest error are displayed, this result could indicate that in January, precipitation level 4 has historically (in the limited sample of cases evaluated) had the highest probability to result in the lowest forecast error. Thus, the results will provide operational hydrologists with guidance as they aim to produce more accurate forecasts. Detailed interpretation of heat maps produced as part of this study is provided in the results section below.

Conventional visualizations (bar plots) were used to show ANOVA analysis results (sum square error).

3. Case Study and Results

The Ohio River Basin covers over 200,000 square miles across 14 states and is populated by more than 27 million people in more than 2400 municipal jurisdictions. As one of the 13 RFCs of the NWS, the Ohio River Forecast Center (OHRFC) is responsible for monitoring of more than 900 streamflow points, 700 of which have real-time flood forecasts. The OHRFC makes use of several models within the CHPS operational environment, including the SAC-SMA model (Burnash and McGuire, 1973; Burnash, 1995), the SNOW-17 snow accumulation and ablation model (Anderson, 1973), several hydrologic routing models, and three reservoir simulation models.

The OHRFC reviews weather and streamflow data daily at hundreds of locations, and updates river forecasts for each. Fig. 3 provides a representative illustration of the OHRFC river forecast hydrograph. The forecasting operation uses previously calibrated hydrologic models combined with weather forecasts to estimate flood hazard potential in the near future. Lead time in this study was defined as the difference between t_{now} and the time of the observed crest. The difference between t_{now} and t_0 is always the same, 30-hours (1-day + 6-hours). The full simulation from t_0 to t_{end} is always 102 h (4-days + 6-hours). That means the forecast period (t_{now} to t_{end}) is 72 h (3-days). Since these are headwater basins that crest within about 30-hours, that time frame captured all the relevant lead times. Six sets of t_0 and t_{now} were created for each crest. The first set is for t_0 at the synoptic time prior to the observed crest, and each subsequent one is six hours earlier. The precipitation modifiers were then applied to the model from t_0 to t_{now}, and the observed precipitation was applied to the period between t_{now} and t_{end}.

While this paper aims to help operational hydrologists better match the magnitude of flood events better through the use of statistical and visualization tools, there are other perspectives on the improvement of flood forecast accuracy. Ehret and Zehe (2011) introduced a new metric “series distance” to quantify the similarity in occurrence, magnitude and time of flood events. Zappa et al. (2013) developed a “peak-box” approach providing visual support that envelops all ensemble peak timings and peak discharge, from which specific verification metrics are

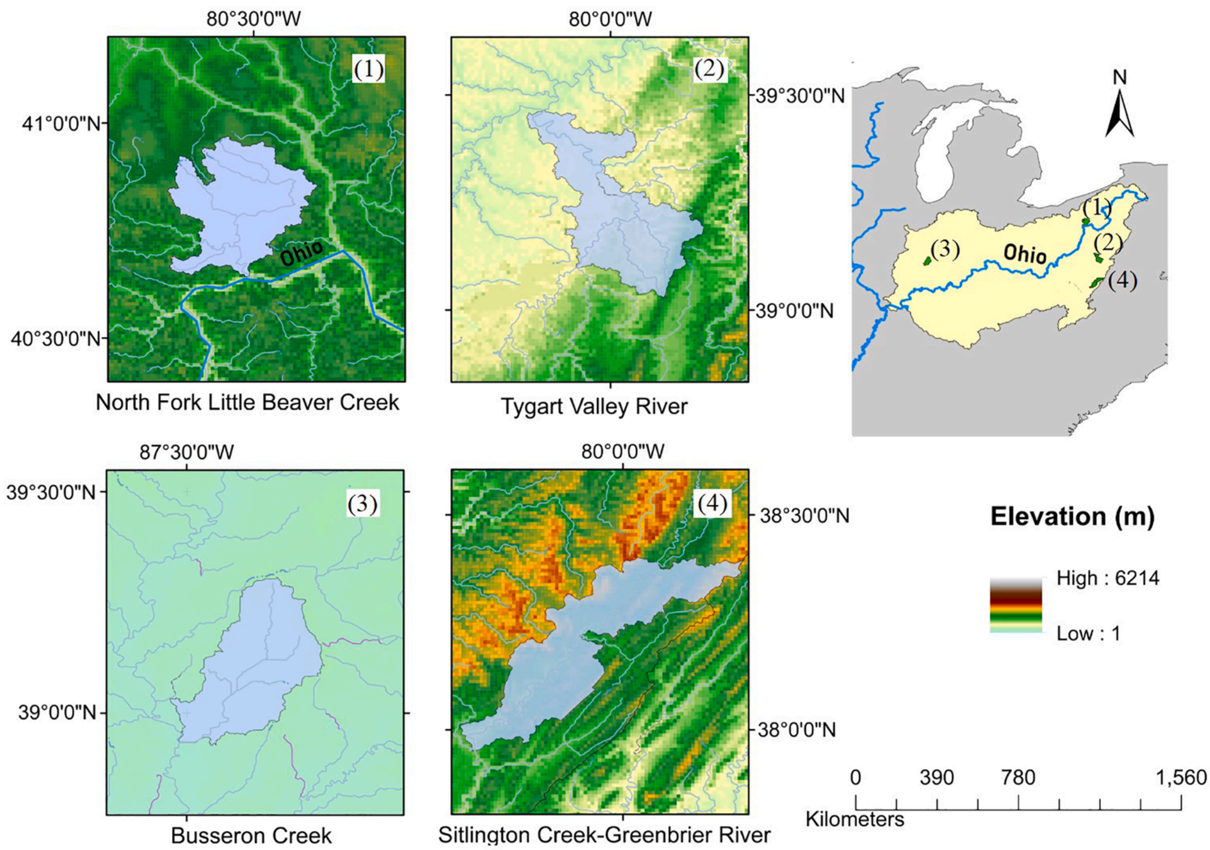

Fig. 4. Case study locations.

defined.

3.1. *Study Area*

The Ohio River Basin is region 5, HUC-2 (among the 21 US Hydrologic Unit Codes). It has 14 HUC-4 sub regions that stretch over the states of Illinois, Indiana, Ohio, Kentucky, Virginia, West Virginia, Pennsylvania, Maryland, Tennessee and North Carolina.

Four HUC-10 watersheds have been selected for case study: (1) Busseron Creek (CRLI3), a watershed in the Kentucky-Indiana area; (2) Sitlington Creek-Greenbrier River (DRBW2), a watershed in West Virginia (3) North Fork Little Beaver Creek (ESTO1), a watershed in the Ohio-Illinois area; (4) Tygart Valley River at Dailey (DLYW2), a watershed in West Virginia. Fig. 4 above shows the location map of the four HUC-10 case study watersheds. Headwaters were selected to minimize the influence of human activities on flood behavior (especially agricultural irrigation extractions and USACE operation of locks and dams), and to eliminate the complexity of channel routing. See Table S1 for detailed information on case study basins.

3.2. *ANOVA*

3.2.1. *ANOVA application*

ANOVA analysis partitioned the uncertainty in OHRFC's 5-day flood forecast. A total of $L = 245$ ensembles were generated for each evaluated historical flood event and lead-time (7 levels for precipitation, 5 levels for soil moisture, and 7 shapes for unit hydrograph). The adjustment to the precipitation is strictly applied for the observed values. When the simulated flow fails to match the observed flow, the precipitation modifiers adjust the "past precipitation", either increasing or decreasing it to produce results as close to the observation as possible at any given time t. The precipitation modifiers were applied to the model from t_0 to t_{now} (see Fig. 3), and the observed precipitation was applied to the period between t_{now} and t_{end}.

Table 3
Specific levels of factors for operational hydrologic forecasts in the Ohio River Valley.

Precip		Soil Moisture		Unit Hydrograph	
1	−90%	1	−50%	1	Baseline
2	−80%	2	−30%	2	Left Shift
3	−50%	3	Baseline	3	Right Shift
4	Baseline (100%)	4	150%	4	High Pass
5	150%	5	200%	5	Low Pass
6	250%			6	Dispersion 1
7	500%			7	Dispersion 2

The initial selection set the magnitude of the change of factor levels (increase/decrease) as a percentage shifts from baseline levels. For precipitation the range of change was ±30%, for soil moisture the range of change was 0–100%, and for the unit hydrograph the levels were: right-left shift, vertical stretch-shrink and horizontal stretch-shrink. The initial trial runs showed flaws in the initial selection that they did not fully capture the magnitude of the flood events (for Busseron Creek, in particular). Thus, a slightly modified selection expanded the factor levels to those shown in Table 3. The Table 3 factor levels provided the necessary resolution for factor adjustments, adequately covered extreme events, and represented settings frequently chosen by the OHRFC operational hydrologists during their regular forecast adjustment practices.

3.2.2. *System to Apply Modifications*

The objective in this study is to guide forecasters in their application of modifiers to improve forecasts. A simple approach to this type of problem is a what-if test (Alberto Benitez-Andrades et al., 2018). A

Table 4
Detailed information of selected flood events.

Month	Number of events each month			
	North Fork Little Beaver Creek	Tygart River Valley at Daily	Busseron Creek	Sitlington
1	47	38	1	0
2	42	38	1	3
3	70	46	2	1
4	54	39	3	0
5	41	41	4	2
6	33	13	2	2
7	23	22	2	0
8	16	5	0	1
9	12	7	0	0
10	14	12	0	0
11	20	18	0	0
12	51	35	4	0
Events time range	1991-4-10 to 2016-12-27	2002-1-12 to 2018-11-16	2008-7-13 to 2015-12-31	2013-5-8 to 2016-6-23
Total events	423	314	19	9

what-if test evaluates the performance of the forecast with and without a modification. One limitation of this method is that the effect of individual modifiers can only be compared to the unmodified case. Operational forecasters routinely apply multiple modifiers, and the modifiers themselves can have a range of values. To find the optimum family of modifiers, all potential modifiers across all potential values would need to be evaluated.

Car manufacturers will perform car crash tests before launching a new model. Likewise, hydrologic model developers usually perform "crash-tests" to ensure the model is safe for use (Andŕeassian et al., 2009). This study applied an expert system approach (Palmer & Holmes, 1988) which elucidates the likely modifiers and their ranges of values from expert knowledge or judgement. A query of the OHRFC database identified the three modifier types most frequently used and the discrete values of each that cover the range of possible values (see Table 4). This led to 245 combinations of modifiers to apply for each hydrologic event for each basin. Perrin et al. (2008) also described a "discrete parameterization" method to calibrate parameters of rainfall-runoff models, which provided a robust parameter set when streamflow time series available for calibration were less than two years.

OHRFC created a system to execute the model at six different lead-times and summarize results under each of the 245 combinations of modifiers and levels. The data extracted from the OHRFC was then processed using computer algorithms written in R language. The algorithm calculated general indexes based on the data sets (such as ensemble identifiers and lead-time of flood forecasts, percent forecast error, etc.) so that the data sets were ready for ANOVA analysis and heat map application. Table 4 provides detailed information of selected events for this study. Only limited data were available for Busseron Creek and Sitlington.

3.2.3. ANOVA Results

Precipitation level 7 (baseline + 500%) dominated the ANOVA analysis, and provided no decision-relevant information. It was therefore excluded from ANOVA results figures. Fig. 5(a) shows the ANOVA results at 0 lead-time, and Fig. 5(b) shows the same results at 30-hour lead-time. ANOVA results shown in Fig. 5 excluded residuals because their inclusion made the distribution results of selected factors less easily comparable. Fig. S2 (supplemental information) shows the same results with residuals included, for comparison. It is clear that a large portion of uncertainty comes from the model itself. A histogram of model error (ensembles without applying "mods") of all flood events at 0-hour lead-time is also provided in the Supplemental Information.

In Fig. 5, the *x*-axis shows factor type, the *y*-axis shows the sum square error of each factor. The λ values chosen for "box-cox" transformation for the four case studies were: 1) 0.28 for 0-hour lead time, 0.2 for 30 h lead time for North Fork Little Beaver Creek, 2) (0.25, 0.24) for Tygart River Valley at Daily, 3) (0.44, 0.2) for Busseron Creek, and 4) (0.19, 0.21) for Sitlington Creek-Greenbrier River. Fig. 5(a) shows that at 0-hour lead time, soil moisture is the dominant uncertainty in North Fork Little Beaver Creek and Busseron Creek basin. The physical

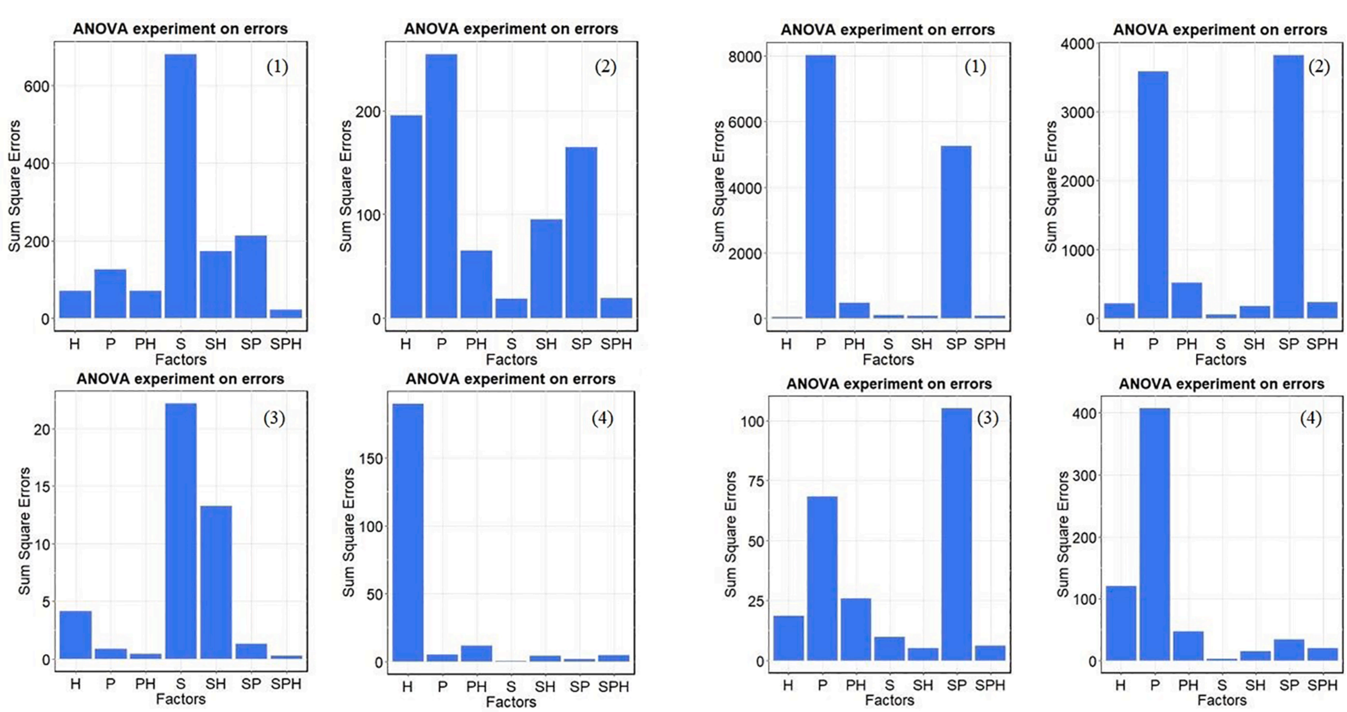

Fig. 5. ANOVA bar plot for factor combinations at (a) 0-hour and (b) 30-hour lead-times: (1) North Fork Little Beaver Creek; (2) Tygart River Valley at Dailey; (3) Busseron Creek; (4) Sitlington Creek-Greenbrier River.

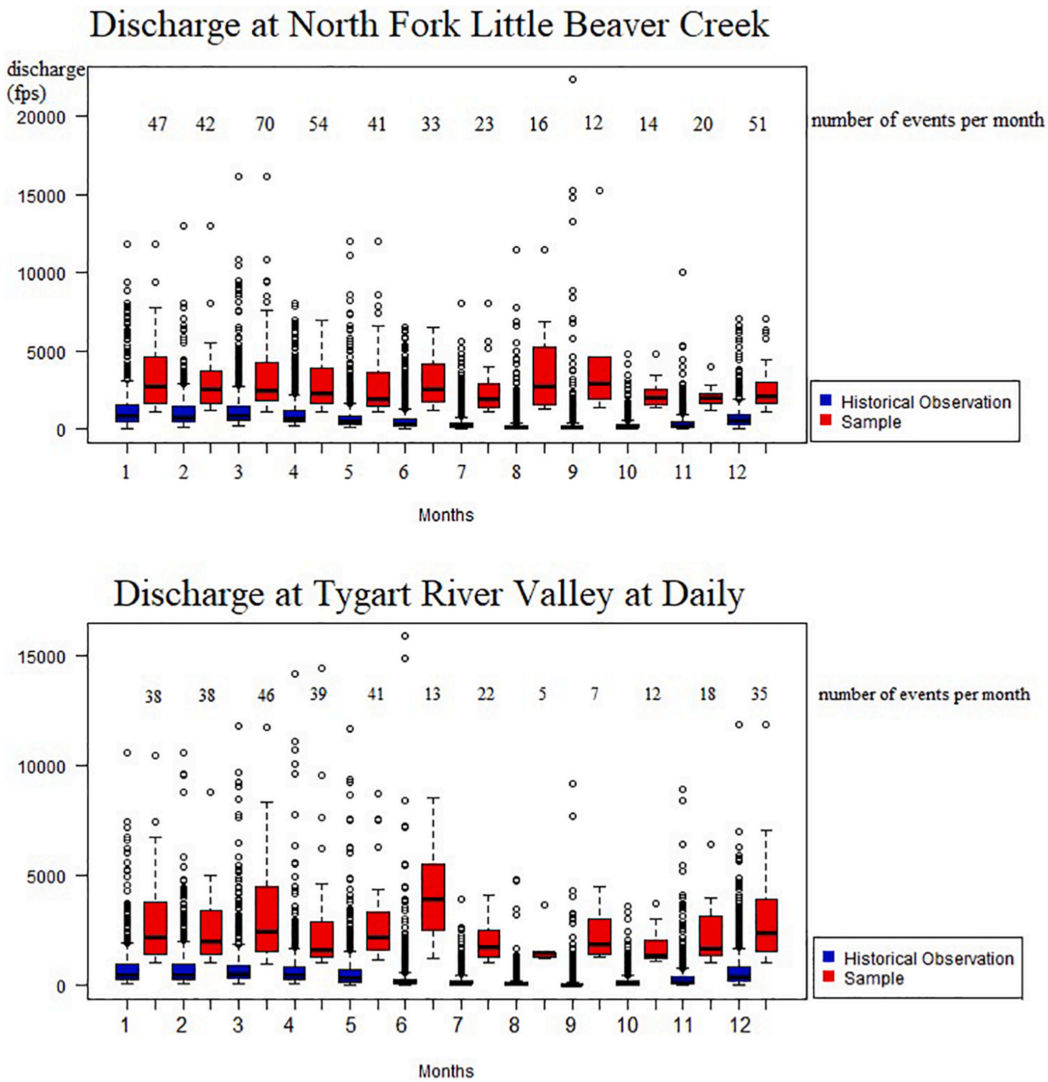

Fig. 6. Discharge of historical observations and sample events at North Fork Little Beaver Creek (a) and Tygart River Valley at Daily (b).

characteristics (see Table S1) of these two basins are highly alike, and they are of similar size (around 600 Km^2) and landcover type (large area covered by cultivated crops). In each of these two basins, agriculture (with direct impact on soil moisture) is prominent. In Tygart Valley River, however, the dominant uncertainty is the precipitation forecast. Tygart Valley River is the smallest in size (479.15 km^2) and mainly covered by deciduous forest. In Sitlington Creek-Greenbrier River, the largest uncertainty source is the unit hydrograph. This is the largest basin (875 km^2), which magnifies the uncertainty in assumptions made regarding the unit hydrograph – uniform basin-wide precipitation, and the occurrence of all precipitation within a pre-defined unit of time. Forecasters should exercise caution in adjusting these factors at short lead-times, as small adjustments in these factors could rapidly increase (or decrease) forecast error.

At 30-hours of lead-time (Fig. 5b), precipitation becomes the largest source of uncertainty in all studied basins. Precipitation is well-known to be one of the most important inputs to hydrologic models but is very difficult to predict with accuracy. The discussion of heat map results in the next section addresses the impact on precipitation uncertainty of seasonality and basin geography.

Because residuals dominate all other sources of uncertainty and make interpretation of the figure difficult, they were removed from ANOVA results, see Supplemental Information Fig. S2 for ANOVA results with residuals.

3.3. Heatmap Results

To emphasize those modifiers that work best in each basin and month (and avoid clutter in visualizations), only combinations of modifiers resulting in the lowest forecast errors are included in heat map results. Only the single level of each of the three modifiers that provided greatest forecast accuracy was selected for each flood event and lead-time (1 precipitation modifier, 1 soil moisture modifier, and 1 unit hydrograph modifier). A description of the procedure for development of the heat maps in this study is provided in the Supplemental Information.

Fig. 6 shows historical discharge (from USGS) and sampled events selected for this study. Note that, for the purposes of this study we have sampled only large flood events in summer, so that in some cases summer values appear higher than winter values. That does not mean that floods in summer are larger than winter floods, in general. For the same reason, the number of events in summer is smaller than that in winter. The number of events sampled per month is shown at top of Fig. 6.

From Fig. 6, the selected events for this study are usually at higher river stage, and intended to capture the historical extreme events. Seasonally, discharge in streams throughout the Ohio River basin is higher in winter (November to April) and lower in summer (May to August).

The panels of Fig. 7 show heat map results for North Fork Little

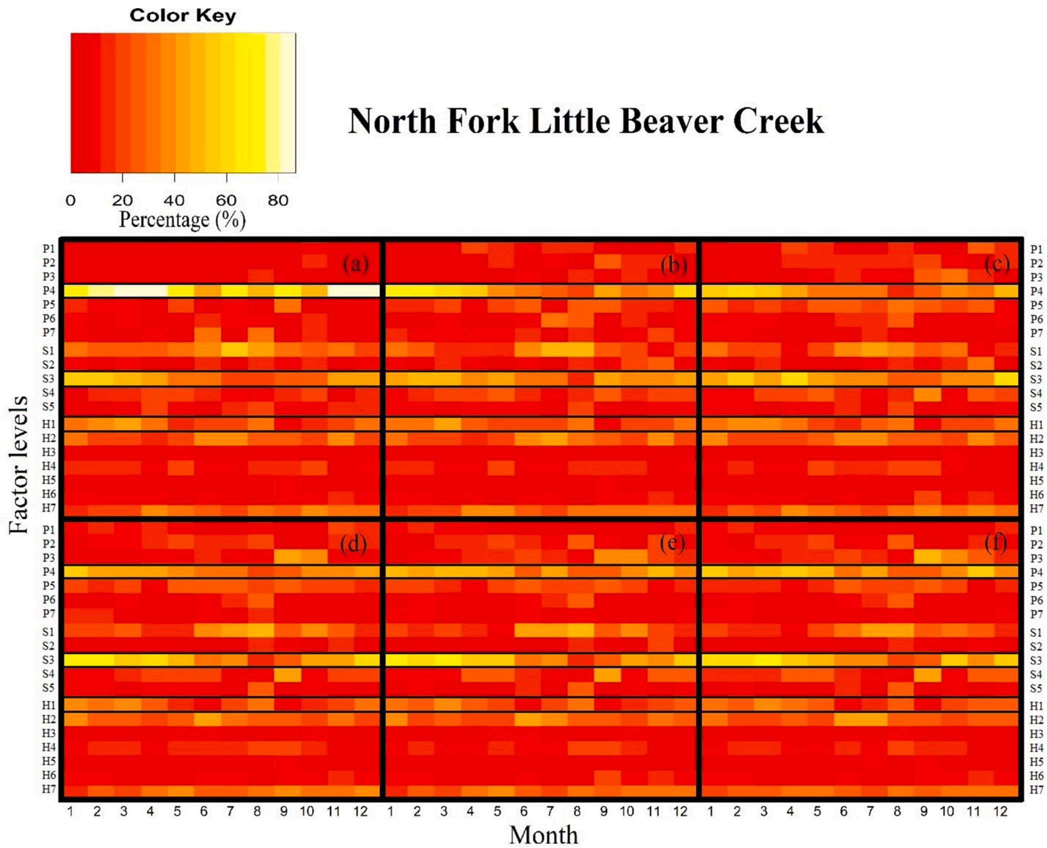

Fig. 7. Heat maps for precipitation, soil moisture and unit hydrograph in North Fork Little Beaver Creek. (a) 0-hour lead-time; (b) 6-hour lead-time; (c) 12-hour lead-time; (d) 18-hour lead-time; (e) 24-hour lead-time; and (f) 30-hour lead-time.

Beaver Creek at each evaluated lead-time: (a) 0-hour; (b) 6-hour; (c) 12-hour; (d) 18-hour; (e) 24-hour; and (f) 30-hour. The *x* axis in each panel represents the month of occurrence for the evaluated historical flood events, and the *y* axis identifies the level of each factor being modified. *P* represents the precipitation modifier and has 7 settings (rows), *S* represents the soil moisture modifier (5 settings/rows) and *U* represents the unit hydrograph modifier (7 settings/rows). The baseline of each modifier is outlined by solid black lines. The color key on the top left shows the percentage value. The redder the color, the smaller the percentage of evaluated historical cases in which this particular modifier setting was included in the best-performing set. Bright colors (yellows/whites) indicate that the modifier setting is the level at which greatest flood forecast accuracy is achieved across a large percentage of the evaluated historical cases.

The lesson from Fig. 7 (North Fork Little Beaver Creek) is: modifications should not be made to the baseline precipitation modifier level in winter; however, alternative precipitation modifier levels may be useful in summer. Focusing on Fig. 7(a), which is the result for a 0-hour lead-time, it is seen that precipitation modifiers have seasonal trends. During winter months (November to April), more than 70% of the best forecast ensembles used baseline precipitation (level 4), which indicates that adjustment in precipitation has not, in these historical cases, increased forecast accuracy during cold seasons and short lead-times. This corresponds with the ANOVA finding, that precipitation is not a dominant uncertainty at short lead-times, and therefore the potential benefits of adjusting it are small. Common sense tells us that the shorter the lead-time, the more accurate the precipitation forecast will be. As the lead-time increases (Fig. 7b-f), although results still show seasonality for precipitation, the percent distribution among levels of precipitation modifiers spreads out, indicating that adjustments to produce better forecasts at longer lead-times might be beneficial, especially during summer months (May-August). This finding could be explained with reference to synoptic and convective storm types. The Ohio River watershed tends to receive synoptic precipitation events in winter months and convective events in summer months (Archambault et al., 2008; Lombardo & Colle, 2011). Synoptic events are easier to forecast, while convective events often result in rapid precipitation and flash floods, which are highly random and difficult to accurately predict. Thus, the considerations of modifier settings representative of deviations from the baseline precipitation forecast are likely to be more fruitful in summer.

For soils moisture, the lesson is: “Don’t touch the baseline soil moisture modifier from baseline except possibly to decrease it to level one (in the event that the recent past in the watershed was very dry).” Level 3 is the baseline. Seasonality in this parameter is seen as well, as it is linked to precipitation seasonality. Thus, forecasters should take care when adjusting soil moisture conditions during winter/summer. However, different from precipitation, a level of the soil moisture modifier (level 1) is a part of the best-performing set often enough that it is a viable alternative to leaving the soils moisture modifier at its baseline level (this is especially during summer months). During hot and dry events common to summer months, the soil moisture may be lower than the default; thus, too much water may have been added to the hydrologic model, and decreasing the soil moisture condition is an efficient way to reduce the error. In that case level 1 of the soil moisture modifier would beneficially be selected. It is very unlikely that departing from baseline soil moisture levels by selecting higher levels of soil moisture would be productive in the summer.

If any adjustments are needed during winter months, the course of action most likely to improve the forecast accuracy is an adjustment to the unit hydrograph modifier. For the unit hydrograph, level 1 (baseline), 2 (left shift) and 7 (dispersion 2) are all frequently members of the

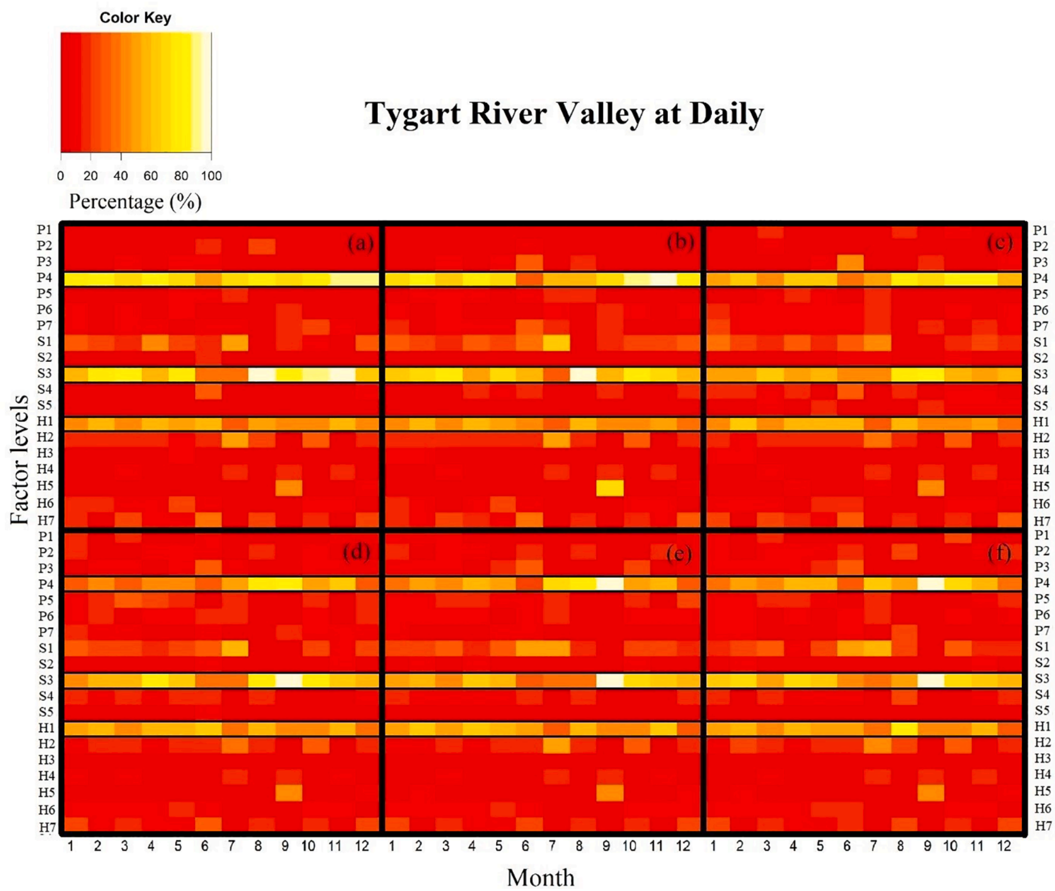

Fig. 8. Heat maps for precipitation, soil moisture and unit hydrograph in Tygart Valley River at Daily. (a) 0-hour lead-time (b) 6-hour lead-time (c) 12-hour lead-time (d) 18-hour lead-time (e) 24-hour lead-time (f) 30-hour lead-time.

best-performing set for this basin. Different from precipitation and soil moisture results, the unit hydrograph modifier is only weakly responsive to seasonality and lead-time. This is a meaningful finding because when simply adjusting precipitation and soil moisture does not provide significant improvement in forecast accuracy, the unit hydrograph could be the factor to be adjusted. Forecasters using the unit hydrograph modifier to improve forecast accuracy should consider performing a left-shift (level 2) to make flood peak sooner, or dispersion (level 7), to add a 5-ordinate smoothing to better match the flood hydrograph.

Fig. 8 shows the heat map results in Tygart Valley River at Dailey. The pattern of precipitation and soil moisture is similar to that for North Fork Little Beaver Creek, except that fewer adjustments to baseline factor levels are constructive across the historical cases evaluated as part of this study. This could indicate a higher accuracy of the precipitation forecast product and/or a better model calibration for this basin. Higher accuracy of flood forecasts resulting from modification of precipitation, soil moisture or unit hydrograph factor levels is unlikely. The best course of action for improvement of forecast accuracy in this basin would be improvement of the model calibration and reduction of residuals (see ANOVA Fig. S2).

4. Discussion and Limitations

This paper presents techniques for the efficient reduction of errors in NWS flood forecasts. ANOVA experiments identified the relative magnitudes of the uncertainty contributed by each uncertain factor. Heat maps illustrate the uncertainty in flood forecasts by season and lead-time for North Fork Little Beaver Creek and Tygart Valley River at Daily. ANOVA results show that the variation from any individual or combination of modifiers is generally an order of magnitude less than the variation from the residuals, indicating the largest source of uncertainty is from the model and its calibration. Residuals aside, ANOVA analysis also shows that soil moisture and unit hydrograph contribute more uncertainty at shorter lead-times, while precipitation is the largest source of uncertainty at longer lead-times.

Heat maps show that at short lead-times, it is nearly always better to use the baseline level for precipitation. Even at large lead-times, changes to the baseline level of the precipitation modifier should not be made in winter; however, alternative precipitation modifier levels may be useful in summer (during localized convective events for which accurate precipitation estimates are elusive). For soil moisture, changes should not be made to the baseline level except possibly to decrease it to level one (in the event that the recent past in the watershed was very dry). Different from precipitation and soil moisture results, the unit hydrograph modifier is only weakly responsive to seasonality and lead-time. Left-shifts (to modifier level 2), or dispersion (level 7), to make flood peak sooner or add a 5-ordinate smoothing, respectfully, have been productive historically in the evaluated basins. Generally, however, none of the changes to modifier values are consistent improvements on baseline levels. It therefore appears to be good policy, except in a few specific circumstances, to leave the modifiers unchanged.

Limitations of this work include: scarcity of information available on NWS RFC historical and current practice regarding the use of modifiers, relatively small study-area basins without channel routing, and a small number of basins studied. Thus, it is unknown if findings of this study apply to downstream basins with channel routing, larger sub-basins, or other climatologic regions. Further study of a variety of basins is needed in order to evaluate the general applicability of the findings presented here.

5. Conclusion and Next Steps

This study shows the performance of ensembles of river flow forecasts generated with different modifier levels. ANOVA analysis suggests largest source of uncertainty comes form the model, thus, better calibration of hydrologic models may reduce overall uncertainty. The results also show that soil moisture and unit hydrograph are more dominant uncertainty sources at short lead times, while precipitation is the primary source of uncertainty at long lead times.

Heat maps generated as part of this study show that both precipitation and soil moisture behavior have seasonality, and that real-time modifications made by NWS operational hydrologists might be useful in the summer season, but are unlikely to be fruitful in the winter. Operational hydrologists can also refer to the heat map to help decide which modifiers and what level to choose when adjustments are needed under specific month and lead-time.

As mentioned earlier, adjustments to precipitation cannot improve forecast accuracy in seasons marked by synoptic precipitation patterns in which precipitation forecast accuracy is already reasonably strong. Opportunities for reduction of precipitation uncertainty in conditions of convective precipitation is warranted.

Systematic collection and analysis of the patterns of use for most-frequently adjusted operational hydrologic modifiers has not been conducted at NWS RFC's, and that is needed in order to improve the usefulness of these findings.

As discussed in Ray et al. (2019), improved operational practices in water systems planning and management hold great promise for improvement to system resilience. If floods could be forecasted more accurately, what would be the resultant reduction in the burden for climate change adaptation measures involving the construction of new infrastructure or the altering of operating procedures? These tradeoffs should be explored.

Finally, the heat map results presented in this paper are probability-neutral. No one combination of uncertain factors is deemed to be more or less likely than another. For example, the combination of modifications "high precipitation/low soil moisture" is given the same likelihood weight as "high precipitation/high soil moisture", when it is clear that one of those combinations is more likely than another. Furthermore, it has been demonstrated in this paper that calibration biases in both the unit hydrograph and the soil moisture estimator result in non-symmetrical probability distributions for each of those inputs to hydrologic forecasts. It may therefore be useful to attempt to characterize the multidimensional uncertainty space for all inputs in combination using hierarchical Bayesian Belief Network techniques such as those described in Taner et al. (2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the University Corporation for Atmospheric Research: COMET Program (National Oceanic and Atmospheric Administration Prime Award Number: NA16NWS4670042, UCAR subcontract SUBAWD000724). We specially acknowledge the contributions of Brian Astifan and the operational hydrologists of the OHRFC, who played an integral role in this study. Finally, this manuscript was greatly improved by the work of two expert reviewers, and we gratefully acknowledge their efforts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.hydroa.2020.100073.

References

Adams, T. E., & Pagano, T. C., 2016. Flood forecasting: A global perspective. Flood forecasting (pp. xxiii-xlix) Elsevier Inc. DOI:10.1016/B978-0-12-801884-2.09999-0.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., Seibert, J., 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. Water Resour. Res. 50 (10), 7541–7562. https://doi.org/10.1002/2014WR015549.

Alberto Benitez-Andrades, J., Rodriguez-Gonzalez, A., Benavides, C., Sanchez-Valdeon, L., Garcia, I., 2018. A semantic social network analysis tool for sensitivity analysis and what-if scenario testing in alcohol consumption studies. Int. J. Environ. Res. Public Health 15 (11), 2420. https://doi.org/10.3390/ijerph15112420.

Anderson, E.A., 1973: National Weather Service River Forecast System—Snow accumulation and ablation model. NOAA Tech. Rep. NWS-HYDRO-17, 229 pp.

Andŕeassian, V., Perrin, C., Berthet, L., Moine, N.L., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.H. and Vaĺery, A., 2009. HESS Opinions "Crash tests for a standardized evaluation of hydrological models". Hydrol. Earth Syst. Sci., 13, 1757–1764, 2009, www.hydrol-earth-syst-sci.net/13/1757/2009/.

Antonetti M. & Zappa M., 2018. How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps. Hydrol. Earth Syst. Sci., 22, 4425–4447, 2018. DOI:10.5194/hess-22-4425-2018.

Archambault, H.M., Bosart, L.F., Keyser, D., Aiyyer, A.R., 2008. Influence of large-scale flow regimes on cool-season precipitation in the northeastern united states. Mon. Weather Rev. 136 (8), 2945–2963. https://doi.org/10.1175/2007MWR2308.1.

Baeck, M.L., Smith, J.A., 1998. Rainfall estimation by the WSR-88D for heavy rainfall events. Weather Forecasting 13 (2), 416–436. https://doi.org/10.1175/1520-0434(1998)0132.0.CO;2.

Boelee, L., Lumbroso, D.M., Samuels, P.G., Cloke, H.L., 2019. Estimation of uncertainty in flood forecasts—a comparison of methods. J. Flood Risk Manage. 12 (S1) https://doi.org/10.1111/jfr3.12516.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resour. Res. 49 (3), 1523–1536. https://doi.org/10.1029/2011WR011533.

Box, G.E.P., 1979. Robustness in the strategy of scientific model building. In: Launer, R. L., Wilkinson, G.N. (Eds.), Robustness in Statistics. Academic Press, pp. 201–236. https://doi.org/10.1016/B978-0-12-438150-6.50018-2.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Stat. Soc.: Ser. B (Methodol.) 26 (2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Brown, J.D., Demargne, J., Seo, D.-J., Liu, Y., 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 (7), 854–872. https://doi.org/10.1016/j.envsoft.2010.01.009.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M.Z., Zhu, Y.J., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. Monthly Weather Rev. 133 (5), 1076–1097. https://doi.org/10.1175/MWR2905.1.

Burnash, R., 1995. The NWS river forecast system—catchment modeling. In: Singh, V. (Ed.), Computer Models of Watershed Hydrology, first ed. Water Resources Publication, pp. 311–366.

Burnash, R. F. & McGuire, R., 1973. A generalized streamflow simulation system: Conceptual modeling for digital computers. NWS Tech. Rep., 204 pp.

Butts, M.B., Falk, A.K., Xuan, Y. & Cluckie, I.D., 2007. Integrating meteorological and uncertainty information in flood forecasting: the FLOODRELIEF project. Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management. IAHS Publ. 313, 2007.

Candille, G., Cote, C., Houtekamer, P.L., Pellerin, G., 2007. Verification of an ensemble prediction system against observations. Monthly Weather Rev. 135 (7), 2688–2699. https://doi.org/10.1175/MWR3414.1.

Deltares, 2018: Flood Early Warning System (FEWS). Deltares, https://www.deltares.nl/en/software/flood-forecasting-systemdelft- fews-2.

Dingman, S. L., 2014. Physical Hydrology. Publisher: Waveland Press, Inc.; 3rd edition (December 17, 2014). ISBN-13: 978-1478611189.

Ehret, U., Zehe, E., 2011. Series distance – an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. Hydrol. Earth Syst. Sci., 15, 877–896, www.hydrol-earth-syst-sci.net/15/877/2011/ DOI:10.5194/hess-15-877-2011.

Fread, D.L., 1973. A dynamic model of stage-discharge relations affected by channeling discharge. NOAA technical memorandum NWS HYDRO-16.

Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P., 2003. Advances in automatic calibration of watershed models. Calibration of watershed models (pp. 9-28). Washington, D. C: American Geophysical Union. DOI:10.1029/WS006p0009.

Habib, E., Qin, L., 2013. Application of a radar-rainfall uncertainty model to the NWS multi-sensor precipitation estimator products. Meteorol. Appl. 20 (3), 276–286. https://doi.org/10.1002/met.301.

IFRC, 2015. World disaster report. Focus on local actors, the key to humanitarian effectiveness.

Janiskova, M., Fielding, M., Crepulja, M., Vasiljevic, D., Kŕal T. and Lean P., 2018. Operational Assimilation of Space-borne Radar and Lidar Cloud Profile Observations for Numerical Weather Prediction. European Centre for Medium-Range Weather Forecasts, ESA ESTEC contract 4000116891/16/NL/LvH.

Kang, T.H., Kim, Y.O., & Hong, I.P. (2010). Comparison of pre- and post-processors for ensemble streamflow prediction. Atmospheric Science Letters, Atmos. Sci. Let. 11: 153–159. doi: 10.1002/asl.276.

Keim, D.A., 2002. Information visualization and visual data mining. IEEE Trans. Visual Comput. Graphics 8 (1), 1–8. https://doi.org/10.1109/2945.981847.

Kitzmiller, D., Miller, D., Fulton, R., Ding, F., 2013. Radar and multisensor precipitation estimation techniques in national weather service hydrologic operations. J. Hydrol. Eng. 18 (2), 133–142. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000523.

Klazura, G.E., Thomale, J.M., Kelly, D.S., Jendrowski, P., 1999. A comparison of NEXRAD WSR-88D radar estimates of rain accumulation with gauge measurements for high- and low-reflectivity horizontal gradient precipitation events. J. Atmos. Oceanic Technol. 16 (11), 1842–1850. https://doi.org/10.1175/1520-0426(1999)0162.0.CO;2.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research, 35(9), pp 2739-2750, 0043-1397/99/1999WR900099509.00.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour. Res. 43 (7) https://doi.org/10.1029/2006WR005756.

Lombardo, K.A., Colle, B.A., 2011. Convective storm structures and ambient conditions associated with severe weather over the northeast united states. Weather Forecasting 26 (6), 940–956. https://doi.org/10.1175/WAF-D-11-00002.1.

Milton, J.S., Arnold, J.C., 1990. Introduction to Probability and Statistics. McGraw-Hill, Singapore.

Najafi, M.R., Moradkhani, H., Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. Hydrol. Process. 25 (18), 2814–2826. https://doi.org/10.1002/hyp.8043.

NOAA, 2010. Office of Hydrologic Development Hydrology Laboratory Strategic Science Plan. DOI:10.1016/S1474-4422(10)70153-8.

NOAA, 2016. Sacramento soil moisture accounting (SAC-SMA) model. Version: OHD-Core-CHPS-4.5.a.

NOAA, 2016. Weather Fatalities 2016, 2017.

Pagano, T.C., Pappenberger, F., Wood, A.W., Ramos, M., Persson, A., Anderson, B., 2016. Automation and human expertise in operational river forecasting. Wiley Interdisciplinary Reviews: Water 3 (5), 692–705. https://doi.org/10.1002/wat2.1163.

Palmer, R.N., Holmes, K.J., 1988. Operational guidance during droughts: expert system approach. J. Water Resour. Plann. Manage. 114 (6), 647–666. https://doi.org/10.1061/(ASCE)0733-9496(1988)114:6(647).

Perrin, C., Andréassian V., Serna, C.R., Mathevet,T. and Moine, N.L., 2008. Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. Water Resources Research, 44(8), DOI:10.1029/2007WR006579.

Pliske, R.M., Crandall, B., Klein, G., 2004. Competence in weather forecasting. In: Smith, K., Shanteau, J., Johnson, P. (Eds.), Psychological Investigations of Competence in Decision Making. Cambridge University Press, Cambridge, pp. 40–70.

Ray, P., and Brown, C., 2015. Confronting Climate Uncertainty in Water Resources Planning and Project Design: The Decision Tree Framework. World Bank Group, Washington, D.C., DOI:10.1596/978-1-4648-0477-9. https://openknowledge.worldbank.org/handle/10986/22544.

Ray, P.A., Taner, M.Ü., Schlef, K.E., Wi, Sungwook, Khan, H.F., Freeman, S.S.G., Brown, C.M., 2019. Growth of the decision tree: advances in bottom-up climate change risk management. J. Am. Water Resour. Assoc., Featured Collection Clim. Change Solutions 55 (4), 920–937. https://doi.org/10.1111/1752-1688.12701.

Roulin, E., Vannitsem, S., 2015. Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. Hydrol. Process. 29, 1434–1449. https://doi.org/10.1002/hyp.10259.

Skitka, L.J., Mosier, K.L., Burdick, M., 1999. Does automation bias decision-making? International Journal of Human - Computer Studies, 51(5), 991-1006. DOI:10.1006/ijhc.1999.0252.

Smith, J.A., Seo, D.J., Baeck, M.L., Hudlow, M.D., 1996. An intercomparison study of NEXRAD precipitation estimates. Water Resour. Res. 32 (7), 2035–2045. https://doi.org/10.1029/96WR00270.

Taner, M.Ü., Ray, P.A., Brown, C.M., 2019. Incorporating multidimensional probabilistic information into robustness-based water systems planning. Water Resour. Res. 55 https://doi.org/10.1029/2018WR022909.

Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. Int. J. River Basin Manage. 6 (2), 123–137. https://doi.org/10.1080/15715124.2008.9635342.

Viessman, W., Lewis, G.L., 2003. Introduction to Hydrology, fifth edition. Pearson Education Inc., Upper Saddle River, NJ.

Wani, O., Beckers, J.V.L., Weerts, A.H., Solomatin, D.P., 2017. Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. Hydrol. Earth Syst. Sci. 21 (4021–4036), 2017. https://doi.org/10.5194/hess-21-4021-2017.

Welles, E., Sorooshian, S., Carter, G., & Olsen, B., 2007. Hydrologic verification - A call for action and collaboration. Bulletin of the American Meteorological Society, 88(4), 50-+. DOI:10.1175/BAMS-88-4-503.

Welles, E., Sorooshian, S., 2009. Scientific verification of deterministic river stage forecasts. J. Hydrometeorol. 10 (2), 507–520. https://doi.org/10.1175/2008JHM1022.1.

Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., Heynert, K., 2013. The delft-FEWS flow forecasting system. Environ. Modell. Software 40, 65–77. https://doi.org/10.1016/j.envsoft.2012.07.010.

Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y.C., Toth, Z., 2008. Ensemble data assimilation with the NCEP Global Forecast System. Monthly Weather Rev. 136 (2), 463–482. https://doi.org/10.1175/2007MWR2018.1.

Wilby, R.L., Dessai, S., 2010. Robust adaptation to climate change. Weather 65 (7), 180–185. https://doi.org/10.1002/wea.543.

Young, C.B., Nelson, B.R., Bradley, A.A., Smith, J.A., Peters-Lidard, C.D., Kruger, A., Baeck, M.L., 1999. An evaluation of NEXRAD precipitation estimates in complex terrain. J. Geophys. Res.-Atmos. 104 (D16), 19691–19703. https://doi.org/10.1029/1999JD900123.

Zalenski, G., Krajewski, W.F., Quintero, F., Restrepo, P., Buan, S., 2017. Analysis of national weather service stage forecast errors. Weather Forecasting 32 (4), 1441–1465. https://doi.org/10.1175/WAF-D-16-0219.1.

Zappa, M., Fundel, F. and Jaun, S., 2013. A 'Peak-Box' approach for supporting interpretation and verification of operational ensemble peak-flow forecasts. Hydrol. Process. 27, 117–131 (2013), DOI: 10.1002/hyp.9521.

Zappa, M., Jaun, S., Germann, U., Walser, A., Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. Atmos. Res. 100 (2), 246–262. https://doi.org/10.1016/j.atmosres.2010.12.005.

Žilinskas, J., Dzemyda, G., & Kurasova, O., 2013. Multidimensional data visualization: Methods and applications. (No.75). New York, NY: Springer. DOI:10.1007/978-1-4419-0236-8.