WIREs Water

Automation and Human Expertise in Operational River Forecasting

Journal:	*WIREs Water*
Manuscript ID	WATER-245
Wiley - Manuscript type:	Overview
Date Submitted by the Author:	26-Jan-2016
Complete List of Authors:	Pagano, Thomas; Bureau of Meteorology, Pappenberger, Florian; ECMWF, Wood, Andy; National Center for Atmospheric Research, Ramos, Maria-Helena; IRSTEA, UR HBAN Persson, Anders ; Sveriges meteorologiska och hydrologiska institut Anderson, Brett; Bureau of Meteorology
Choose 1-3 topics to categorize your article:	

WIREs Wiley Interdisciplinary Reviews

Article Title: Automation and Human Expertise in Operational River Forecasting

Article Type:

☐ OPINION ☒ OVERVIEW ☐ FOCUS ARTICLE

☐ PRIMER ☐ ADVANCED REVIEW ☐ SOFTWARE FOCUS

Authors:

First author Thomas C. Pagano* Bureau of Meteorology floodwarning@bom.gov.au (email to be included in publication) tompagan@bom.gov.au (correspondence with editors)
Second author Florian Pappenberger European Centre for Medium-Range Weather Forecast
Third author Andrew W. Wood National Center for Atmospheric Research
Fourth Author Maria-Helena Ramos Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA)
Anders Persson Swedish Meteorological and Hydrological Institute (SMHI) – Retired
Brett Anderson Bureau of Meteorology

Abstract

Increased automation is an attractive option for hydrologic forecasting agencies faced with growing product complexity and institutional resourcing pressures. Although the hydrologic literature has been nearly silent on the roles of expertise and automation in forecasting, other disciplines such as meteorology have had decades of open discussion on the topic. To address the lack of dialogue in hydrology on automation, this article seeks to contextualize relevant findings from similar disciplines, including meteorology, psychology, decision support systems and interface design. We predict which aspects of operational hydrology have the greatest chance for success at

implementing automation in the near future. Some applications have employed higher levels of automation, notably flash flood forecasting which requires rapid response times, and extended prediction which emphasizes uncertainty quantification. Short-range flood forecasting may be more challenging to automate and traditionally has been less automated than other types of forecasts, partly because of existing practices of interfacing with meteorologists and water system operators, and the difficulties in modelling interference in the water cycle. Overall, we suggest that the design of automated forecasting systems should consider three factors:

1. Processes change under automation and people may require new roles.

2. Automation changes the way people behave, sometimes negatively.

3. People may not have accurate perceptions of the quality of the automated guidance.

Seven lessons learned from automation in meteorology are highlighted and translated into a hydrologic forecasting context, leading to a set of recommendations for how to make best use of expertise in increasingly automated systems.

1. Introduction

Hydrologists strive to provide reliable operational river forecasts that facilitate effective water management and emergency flood protection. Shifting institutional resources and growing complexity – such as an increasing number of data sources and forecasting models, and demand for new forecast products – creates pressure to re-shape hydrologists' involvement with forecast production. Increased automation is one way to increase efficiency, accelerate information generation, and broaden the capacity of forecast centers. Automation enables implementation of advanced techniques that may be inconsistent or incompatible with the traditional manual forecasting paradigm. For example, ensemble forecasting systems deal with more data/models than deterministic systems[1]. Objective data assimilation and streamflow post-processing procedures[2] require a consistent, repeatable process for a statistically robust implementation.

It is a widely held view that experts' contributions add value to warnings and information to stakeholders. If so, increased automation should be accompanied by measures that continue to best utilise forecasters' talents[3]. To the authors' knowledge, there are no systematic studies in the hydrologic research literature investigating the role of forecasters. However, based on research from other fields, there is ample evidence that people have subjective expertise that allows them to consistently outperform objective algorithms in certain contexts[4].

Studies in the climate domain have shown nonetheless, people have cognitive biases that can interfere with the generation and interpretation of forecasts[5]. Manual forecasting is non-repeatable, may lack transparency, and is more difficult to evaluate than automated forecasting. Researchers warn of problems that can arise when people and machines work together, such as the tendency for people to put too much trust in model outputs[6] and difficulties for people to regain control during automation failures [7]. Awareness, training and appropriate system design can limit some of these negative aspects[8].

Despite the lack of hydrologic studies in this area, the role of the forecaster is an active topic of discussion in the meteorological community [9-11]. For instance, Canada's replacement of many of its human forecasters with an automated weather forecasting system raised questions such as: "*If routine weather forecasts are relegated to machines, how can algorithms also alert forecasters for the potential for high impact weather, prompting the human to do more detailed analysis?"* [12, 13]. In meteorology, the availability of supercomputers, widespread use of data assimilation, and an increasing emphasis on probabilistic and ensemble forecasts add to the practical difficulty of adjusting and editing the large volumes of automatically generated forecast information. This makes automation more attractive. However, the meteorological community has recommended forecasters should be sceptical of and critically evaluate model guidance when developing public warnings[9].

While the experience in meteorology is useful to hydrology, the role of expertise in hydrologic forecasting deserves its own discussion. Hydrologists are faced with many challenges that meteorologists do not have to contend with, such as human interference in the water cycle (e.g. reservoirs, irrigation, flood control measures) and the space-time dynamics of watersheds. Discussions about automation have occurred internally at some operational river forecasting centers but have been largely absent from the literature. Questions remain, such as: Aside from the traditional manual practice, what other strategies are viable for applying forecaster expertise to create river forecasts and warnings? On what tasks and situations should a forecaster's efforts be focused and which be automated? Should automation of hydrologic forecasting be a goal?

This article aims to investigate these issues and open a discourse among operational forecasters and researchers on the roles of expertise and automation in river forecasting. The article begins with reviews of the tasks of hydrologists and the state of automation in forecasting (section 2). The main scientific contribution of this article is the synthesis of relevant findings in similar disciplines (section 3) to create predictions of which aspects of operational hydrology have the greatest chance for success at implementing automation in the near future (section 4). Section 4 develops a set of recommendations for making best use of forecaster expertise. The article finishes with a summary of the findings.

2. Operational River Forecasting

2.1 Main operational tasks

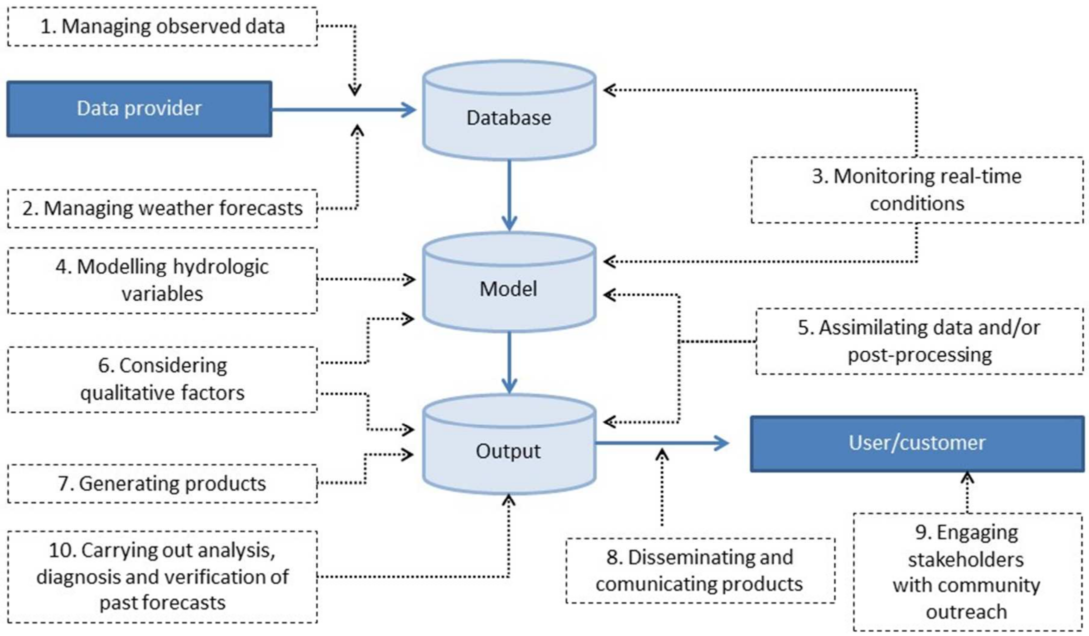


Figure 1: The main operational tasks of forecasters.

Figure 1 synthesizes the operational tasks of a river forecaster. The importance and details of each task depends largely on the context and duties of the service. Detailed descriptions of these tasks can be found in Sene [14], while the focus hereafter is on the main aspects of each task that can play a role in automation. Each task may be done by an individual or shared among personnel. Additionally, one individual may do all tasks or work as a specialist within a group. For example, some agencies distinguish modellers (whose objective is to generate quantitative predictions) from flood warning hydrologists (who synthesize guidance and communicate to users). This article uses "river forecaster" as a generic term for those involved with one or more of the tasks described below.

Although data collection and transmission is largely automated (figure 1, tasks 1 to 3), the river forecaster nevertheless spends non-trivial amounts of time checking, cleaning, infilling, using, archiving and redistributing hydroclimatic data. Hydrologists often visually inspect data to assess its quality but also to recognize if any immediate action is necessary (e.g. a reservoir has just filled and thus a flood warning is necessary for those downstream), or a data provider must be notified that a gauge is offline.

Next, prognosis follows in two stages: The forecasting of future weather conditions and the modelling of hydrology (figure 1, tasks 2 to 5). River forecast accuracy is particularly vulnerable to precipitation forecast displacement and magnitude errors, and (where snow is present) to temperature errors. Due to the large uncertainties in future precipitation, particularly for extreme events, hydrometeorologists may further localize weather forecasts, create contingency scenarios (e.g. rainfall falling in or outside the watershed) or to translate them to another spatial or temporal format.

Critically distinct from meteorology's Numerical Weather Prediction (NWP) models, operational hydrologic models are typically parsimonious and simple – some run in seconds - allowing hydrologists to run them iteratively, with real-time adjustments of their parameters and inputs as events unfold. Forecasters may alter the raw hydrologic model output if compelling anecdotal evidence is available that suggests that the forecast is deficient (figure 1, task 6). For example, rainfall-runoff transformation may be driven by a basin process that is known to be important but is difficult to quantify. The hydrologist may also need to consider non-stationarities (e.g. major changes in land cover following fires) and human factors (e.g. the drop in river height when levees fail).

Forecast formulation is similar to the well-studied process applied by weather forecasters [9, 13, 15, 16]. Forecasters make interpretations and, among other things, try to increase forecast consistency through temporal and spatial smoothing of the model outputs. Smoothed outputs may be less accurate statistically but users usually prefer forecasts that do not waffle[17] e.g. "it will flood", "it will not flood", "it will flood", [17]. The final forecast and contextual data are packaged into textual and graphical forecast products that may include narrative discussion about predicted conditions (figure 1, tasks 7 and 8). The products may also take the form of targeted warnings, such as flood warnings, which may include predictions as well as instructions for remaining safe.

The final stage of disseminating and communicating products may involve operational data exchanges and decision support for consumers and interaction with the media. Hydrologists may also engage stakeholders with community outreach (figure 1, tasks 8 and 9) to raise awareness about, and trust in, the forecasts, but also to help the forecasters better understand the users' needs. In particular, some users struggle with probabilistic forecasts, in part because the concepts are technically complex, but also because their use is more effective when coupled with risk-based decision frameworks, which may not be easily articulated or formalized[18]. Part of the demonstration of the value of the forecasts involves verifying past forecasts against observations (figure 1, task 10)[3,]

[19, 20]. Forecasters may write “post-mortem” evaluations of past significant events[21] generating reports such as the NWS “Service Assessments” (http://www.nws.noaa.gov/om/assessments/).

2.2 Current status of automation

Although it is difficult to generalize about the status of automation in forecasting enterprises, investigating a few key systems can be illustrative[22]. Pagano[23] compared and contrasted the roles of automation in 19 forecasting systems in several developed countries. Pagano encountered three primary modelling paradigms: 1) **passive systems** in which the model is run and products are generated without human adjustment, 2) **observant systems** where people supervise the model and mainly use it as a decision support tool and 3) **engaged systems** where people actively use their expertise in real-time to adjust and, in theory, improve the model runs.

Some river forecasting systems are almost completely manual, such as some early warning systems in developing countries. In Nepal, when river levels cross a threshold, a person uses a hand-cranked siren to alert communities downstream[24]. Here, the forecast skill comes from the delay between upstream and downstream peak flows. The operator relies on a standard operating procedure and hydrologic judgements are unnecessary. This task would be automated but for the relatively low cost of employment in developing countries and the limited, and possibly unreliable, communications infrastructure, specifically during a flood event.

Among the countries that use computer models for forecasting, the US National Weather Service (NWS) has a hands-on engaged forecasting paradigm where the hydrologist is “in the loop”. The process is semi-manual, having evolved to correct for the system’s many data, modelling and science challenges[25, 26]. Hydrologists perform data management, real-time monitoring, manually develop precipitation forecasts, actively manage the forecast model forcing inputs, and manually manipulate states and parameters of hydrologic model and its output. The focus for flooding is on the single-valued flow and stage predictions. Hydrologists may generate products and interact with stakeholders[22], although this task is led by local weather forecast offices in affected areas.

Less well known than these semi-manual official flood forecasts, US agencies have created largely automated and rapidly updated “guidance” products directly from their models[27]. Examples include the daily updating of seasonal streamflow forecasts from statistical models[28] and flood forecasts from dynamical rainfall-runoff models forced with ensemble NWP outputs[3, 29]. The latter set relies on the manual model state maintenance process described earlier, but avoids the real-time modification of forecast meteorological inputs and streamflow outputs. All of the above products are available to the public.

In other countries, national level forecasting often follows a more observant paradigm with a higher level of automation. In the United Kingdom Flood Forecasting Centre (FFC), the hydrologic modelling system creates national gridded maps of flood probabilities, as well as time series at certain locations. Although the hydrographs are considered physically realistic, the real-time predictions are often couched in model climatology exceedences[30]. Data assimilation is automated and the hydrologist mainly interprets the model output to aid in the creation of categorical flood guidance maps and text-based products explaining the situation[31]. A significant part of the work involves coordinating with regional forecasters (who run their own models and have their own perspectives) and liaising with users. FFC share the same systems as the regional forecasters and so can run the same localized models in order to develop a better understanding of forecast flood risks. In contrast to the gridded model output, the FFC's flood guidance statements are impact-based and relate to general flood risk level by county. The public cannot access the model output directly. The flood warnings issued by regional forecasters are available to the public and emergency responders. The FFC system is similar to those operated in France and the Netherlands (Jan Verkade, personal communication 10 June 2014).

Emerging systems having a transnational or global extent typically employ very high levels of automation. The European Flood Awareness System (EFAS[32]) and fledgling global offshoot (GloFAS[33]) are examples of observant systems, though their very high levels of automation make them nearly passive. EFAS is the result of interagency development, primarily led by the Joint Research Centre of the European Commission. Model-running now resides at the European Center for Medium Range Weather Forecasts (ECMWF). Multiple ensemble and deterministic weather forecasts are used as input to EFAS. The outputs are 6-hour to daily streamflows with lead-times up to 15 days ahead. EFAS performs automated streamflow data assimilation[34] at a few dozen points. Hydrologists' responsibilities include monitoring the system running and delivering forecasts to another center responsible for the dissemination of products. Users are forecasters in national hydrological services, since EFAS and GloFAS products are not available to the public. EFAS is successful in its approach because of its data-modelling consistency. Specifically, the hydrology model is forced with real-time NWP ensemble forecasts that are entirely consistent with NWP ensemble hindcasts. Similarly, the hydrology model is operationally initialized with the same data used to generate the model climatology. Streamflows for each pixel in the model domain can be calculated and forecasts are compared to flood thresholds of given return intervals derived from the model climatology [32]. Additional external information is still needed to translate these predicted risk levels into public warnings of local hazards (e.g. which shopping centers will be inundated). Such localized public warnings are the responsibility of national forecasting services that receive EFAS alerts. National services often use these alerts as a "heads-up" that flooding is possible and base the warnings on further analysis using in-house tools.

Finally, the highest level of automation can be found in systems operated at universities and research centers. The University of Oklahoma/NASA provide fully automated flood predictions based on satellite rainfall estimates, NWP outputs, and land surface model simulations[35]. Converting these generalized forecasts into actionable warnings would still require local flood vulnerability information. Such examples of near or fully automated, passive systems suggests that, at least at this stage in their development, they play, at most, a complementary (versus replacement) role to national-scale or regional, engaged flood warning services.

3. Human-Machine Interactions

Many research publications address automation and human-machine interactions. These include studies of psychology, decision support systems and interface design. Experimental evidence comes from the laboratory and the field across professions including doctors, pilots, and judges. This section analyses the research that is most relevant to expertise in river forecasting systems.

3.1 Capabilities and limitations of people and machines

Machines are better at repetitive/routine tasks, applying logic, and multi-tasking. Machines are fast, reliably follow instructions, are consistent, have sustained performance, and their behavior is reproducible. People are better at improvisation, inductive reasoning, and interactions with customers[36]. People are commonly cited as being better at "the big picture", and machines, "the details"[37].

Essentially, machines have logic but lack sense. However, how good are people at these higher cognitive functions? In order to have value in the active involvement of hydrologists in the forecasting process, there must be evidence that people are capable of making intuitive judgments about impending floods. Such skilled intuition is the subjective ability to make accurate sense of a situation, through rapid assessment of environmental factors, and recommend an optimal course of action[4].

Kahneman and Klein[4] synthesised competing schools of thought on the quality of intuitive judgement. Kahneman studied cases in which human judgment was flawed, whilst Klein focused on cases where people recognised the best decision in highly complex situations. Those authors concluded a “high validity” environment is a necessary though insufficient condition for the development of skilled intuition. Such environments present “stable relationships between objectively identifiable cues and subsequent events or between cues and the outcomes of possible actions.” Validity and uncertainty are not incompatible and they cited poker as a valid yet uncertain example of where the best moves reliably increase the potential for success. Unfortunately, high subjective confidence is not a good indication of validity. People also struggle with recognising randomness. Streaks can occur in randomly generated sequences but people too commonly assess streaks as non-random[38].

Finally, algorithms outperform people in low-validity environments since algorithms can identify weakly valid cues and use them more consistently than people. In these cases, statistical models often outperform humans. Models of the judges even outperform the judges themselves, partly due to human inconsistency[39]. It is a challenge to avoid over-fitting models when cues are weakly valid- some of the cues will be spuriously significant and there is a desire for positive outcome. An additional use for models in weakly valid environments is to inform the human which cues are invalid and this should lead to a search for better cues.

Nicholls[5] describes ten cognitive traps climate forecasters and users can fall into: The framing effect; Availability; Anchoring and adjustment; Underweighting base rates; Overconfidence; Added information bias; Inconsistent intuition; Hindsight and confirmation bias; Belief persistence; Group conformity/decision regret. Overconfidence has been called the most pervasive and potent bias to which human judgement is vulnerable. For example, when asked to provide a 90% confidence interval for an estimate of a particular number, people typically give too narrow a range (e.g. one that contains the truth 30% of the time), indicating overconfidence[40]. Over-precision is also remarkably robust and resistant to de-biasing[41]. People are overly optimistic about personal risks, believing hazards are more likely to happen to others than themselves.

Murphy[42] highlighted the possible discrepancy between forecasters’ best judgment and their issued forecasts. Ideally the two should be identical, however, in a hydrological context, for instance, the forecaster may issue a hydrograph forecast with an unreasonable recession rate. They may not truly believe that recession may occur, but their primary goal was to issue an accurate peak forecast and the software limited the ability to satisfy both objectives. This would be a case of inappropriate human-machine interaction. There are cases where forecasters would purposefully issue a forecast that is too high or too low so as to inspire or hedge against action by users, or to smooth out forecast “waffles”. Forecasters are more vulnerable to external and societal pressures than automated systems. Conversely, it could be argued that forecasters are trying to satisfy users’ presumed "holistic" needs, whereas the automated product's sole objective may be maximizing a narrowly defined measure of forecast accuracy.

3.2 Effective Design of Automated Systems

Generally, a few conditions are necessary to successfully delegate tasks to an automated algorithm. Kahneman and Klein[4] said there must be

"1. confidence in the adequacy of the list of variables that will be used,

2. a reliable and measurable criterion [performance measures],

3. a body of similar cases,

4. a cost/benefit ratio that warrants the investment in the algorithmic approach, and

5. a low likelihood that changing conditions will render the algorithm obsolete."

It is best to automate tasks of information acquisition and analysis but people should be able to recognize when automation has gone awry and override automation[43]. When increasing automation in a decision support system, the literature cautions against three issues:

1. Processes change under automation and people may require new roles. According to Dawes[44] people are much better at selecting cues to be considered in a model than they are at integrating the cues. People are also skilled at providing a "sanity check" on the model, such as recognizing when it is relying on bad data or basing its predictions on outliers. Automation can compensate for, or mitigate, the unintended consequences of cognitive bias. Similarly, human supervision can reduce the likelihood of computer-generated errors, misguided predictions and automation failures. The two components – person and machine – can be complementary in a well-designed system, and can extend the human's capabilities.

However, automation will rarely mimic exactly the manual procedures it replaces. If system developers simply pick the most easily automated tasks and replace those first, people are often given "leftover" tasks that may not suit the forecaster's capabilities. An automated system can also present hazards, which can be a large concern if the system is critical to a high-stakes mission. If a skilled operator is decoupled from the workings of the process they are supervising, they may become de-skilled and unable to take over when automation fails[8].

In a hydrologic setting, this means that because forecasters often have very good mental models of how nature behaves, they should work closely with developers to build and implement numerical models that take into account forecasters' knowledge. Also, before automation, forecasters may have a varied set of responsibilities that enriched their experience and improved their mental models, such as cleaning data, executing models, interpreting model output and generating products. If some of these tasks are automated, the remaining tasks may seem monotonous. This can de-motivate forecasters who may spend their time discrediting the automated system, instead of using their expertise to enhance it.

2. Automation changes the way people behave, sometimes negatively. Without vigilance, automation causes problems of mistrust and complacency, degraded situational awareness, and problems with reclaiming control[36]. Skitka *et al.*[6] suggest that under automated conditions, the main problem is no longer operator error, but rather designer error. Furthermore, operator errors still occur, just in a different form. In contrast to the maxim "Garbage In, Garbage Out", the phrase "Garbage In, Gospel Out" describes human over-reliance on automated decision aids.

Doswell[45] suggests that this bias is not just due to a cognitive blind spot, it also relates to personal risk assessment. If an automated system warns of an event and the person chooses to ignore it, they expose themselves to liability and professional risk if the event actually occurs. Conversely, if they issue what they think is likely a false alarm, the repercussions are diffuse.

With automation, hydrologists may be more likely to issue a warning if the model predicts a flood, even if it disagrees with the raw data. The reverse is also true - when the automated warning is potentially present, but silent, the forecaster could do nothing, regardless of what all other indicators suggest should be done[46].

3. People may not have accurate perceptions of the quality of the automated guidance. While people often comply with model suggestions, people also underestimate model output quality. When pilots used a faulty decision support system [6], their subjective impressions of the reliability of the system (e.g. 82% reliable) were worse than the actual (94%). Put another way, people think models are worse than they actually are but still use them anyway. This is most challenging when quality is variable[47], partly because trust is conditioned on the worst outcomes (i.e. largest errors in recent memory [48]). Institutional factors also affect the acceptance of automated guidance. Early performance of a system leaves lasting impacts on operator acceptance, and internal gossip can distort perceptions of a system's capabilities[49].

Forecast verification can be used to ground hydrologists' understanding of automated product performance, especially when compared to a baseline like manually produced forecasts [50]. However, a prototype automated forecasting system may initially perform poorly and could leave the hydrologist with an enduring negative impression (even if errors were atypical and subsequently improved). This hydrologist may even warn colleagues against accepting the system. Therefore, care should be taken to evaluate prototype systems critically, but not in a way that undermines their later potential adoption.

3.3 Relevant lessons from meteorology

Decades ago, NWP models were few enough that a meteorologist could gain familiarity with their tendencies and compensate for failings in the real-time forecasts. Today, forecasters cannot possibly have the same understanding of the dozens of real-time models, thus the traditional manual approach has ceded some ground to semi-objective consolidations and corrections of models[51]. For nearly as long as computer weather models have existed, there has been the suggestion that someday meteorologists will be unable to outperform the NWP models. The warning is of a "meteorological cancer"[52] in which forecasters rely on models unquestionably, atrophy their independent talents and find it difficult to compete. This ultimately leads to forecaster obsolescence.

Researchers recommend seven best practices for meteorologists and system developers:

1. *Use automation to quality-control and ingest data:* Aside from the effective use of high performance computing, meteorology's greatest technological achievement lies in the implementation of automated data assimilation. Weather modellers routinely objectively assimilate tens of millions of four-dimensional observations per day into models with 10^8 degrees of freedom.

2. *Use well-designed forecasting interfaces:* Some studies of meteorological automation focuses on workstations, the primary tool for forecast creation[53]. Meteorology has pioneered the development of Interactive Forecast Preparation software, such as the Graphical Forecast Editor[54]. Here, instead of manually crafting narrative and products, the forecaster interactively edits a set of NWP forecast grids and products are automatically generated from the result. This does not reduce the subjective input to the forecast and allows new, more detailed, products by streamlining the integration task. GFE relies on an underlying digital weather database to ensure that internal consistency and physical realism are maintained even after the forecast has been edited.

3. *Have transparent systems:* To effectively supervise and intervene in automated systems, people need the option to view inputs and intermediate states to determine if the

automated output disagrees with what would be expected. This includes being able to view model output before statistical post-processing is applied, making it easier to diagnose potential errors.

4. *No peeking at the answer:* Meteorologists recommend the separation of prognosis from diagnosis [55]. Although the practice is rare nowadays due to automated chart drawing, Roebber *et al*. [56] recommend that the forecaster should also hand-draw weather charts before looking at the weather model output. This prevents being prejudiced by model output and places meteorologists in a better position to understand and question the model guidance. Occasionally turning off "auto-pilot" during typical conditions keeps up operator training in case of system failure. Experimental evidence consistently shows that forecasters generate considerably better short-range predictions when model guidance is initially withheld and they are forced to spend more time on analyses, diagnoses, and creating their own prognoses [57].

5. *Verify your forecasts:* Rapid, relevant and unambiguous feedback is the key to improving intuitive expertise[4]. Structured forecast evaluation is also critical for directing investments in system improvement and recognizing existing weak spots. In forecast verification, one should avoid viewing evaluators as "the forecast police" or using highly aggregated skill scores. Forecast verification should be stratified to focus on "high impact" and/or difficult forecasts, and be done in an informative way[12].

6. *Never stop learning the science:* To develop expertise, the forecaster must learn to recognize reliable, relevant cues from the environment and be able to respond effectively. Recognition can come from training and experience. Nearly all publications stress the need for better forecaster education and training. Doswell[58] recognized the challenges of operational learning:

"*Instead of having the chance to learn forecasting by doing it, one quickly discovers that the forecasting world is a lousy place for learning. In the rush to get products out, there are few opportunities for leisurely consideration of the meteorological issues. If one makes a bad forecast, there are few opportunities to go back and see what could have been done to avoid that problem.*"

7. *Redefine your role:* The meteorological community is divided on the role of forecasters in decision support. Many agree that the most important task is to help end-users, such as regional and national authorities, to make the optimal decision about protective action[16]. This may involve interacting with customers, transitioning forecasters into the role of communicator and interpreter, and taking some meteorologists away from basic forecast construction[10, 12]. However, this distances the forecaster from the creation of forecasts, potentially limiting the ability to understand and justify it. Furthermore, increased emphasis on "adding value" for users may put government forecasters in competition with those from the private sector.

4. Discussion

Here, we drew on some of the recommendations in meteorology and other fields, framed them in the broader literature on automation and expertise, and translated them to a river forecasting context. The review presented herein supports several clear, high-level messages. Specifically, if the environment has reliable, relevant and observable cues people can use to improve forecasts, they should be given those tools and opportunities. If a process can be relegated to an algorithm, do so, provided people may still supervise and intervene where applicable. The effectiveness of these runtime interventions should be monitored and reported by the forecasters to assist improvement. The automation should lead to synergies between people and machines. Turning the person into a disinterested machine minder should be avoided.

Given the conditions for success discussed in this paper and in the literature, these are candidates for successful automation in hydrology:

1. **Seasonal forecasts** are infrequently issued, and, typically, there is a long delay (one or several months) between forecasts and the outcome, making the process of receiving feedback (e.g. to improve mental models) slow. Also, the relative skill of the forecasts is currently usually low[59] and the forecaster's interventions are often within the limits of the typical errors of the models; This means that the hydrologist will usually not receive definitive proof as to whether the hypotheses used were correct.
2. **Flash flood forecasts** [60] have response times that are very rapid. These are often not based on a formal forecasting model, but instead on nowcasting techniques, issuing alerts based on recent observations and high rates of change. Real-time data must be processed quickly and there is a narrow timeframe to alert the user. Hydrologists would struggle to

provide around-the-clock rapid response without additional resources and staff. Depending on the climate, flash floods may be rare, leaving people without learning opportunities on most days.

3. **Extended and medium-range forecasting** has considerable uncertainty and there is thus greater emphasis on forecast ensembles and quantifying uncertainty. Aside from the workload of intervening in data-rich ensemble products, people are poor at subjectively estimating probabilistic forecast distributions. Simulation models typically exhibit overconfidence because one or more sources of uncertainty are ignored. Post-processing hydrologic models outputs is often necessary [61, 62]. Several techniques exists to make hydrological ensemble forecasts probabilistically reliable [63, 64 65]. The challenge are how to transfer such techniques to operational environments and how to supplement the results with the forecasters' views [66].

For the above systems, there is still a major role for people as system designers, monitors, and interpreters, rather than as "in-the-loop" operators.

In contrast, **short-range riverine flood forecasting** may be more difficult to automate and traditionally has been less automated than other types of forecasts [67 23]. These systems have often been developed based on single-valued forecasts and at local scales, with spatially lumped and parsimonious models. Provided that the correct systems and training are in place, hydrologists can frequently receive rapid and unequivocal feedback when verification is performed, quickly correcting and learning from mistakes. There are also important but difficult to numerically model factors, such as obstructions to flow (e.g. blocked drains), structure failures, and human regulations, each providing opportunities for people to manually enhance the forecasting process. In addition, some catchments are difficult to model because of their extreme climate and/or unusual hydrologic processes.

When making predictions that extend beyond the response time of their river catchments, hydrologists must consider weather forecasts. There are several successful examples of river forecasting systems that automatically use NWP model output [68], provided that care is taken to convert the weather model output into the appropriate form for the hydrology model. Although this enriches the flow forecasting system, the value of the meteorologists' expertise can be lost if a dialogue or communication platform does not exist between meteorologists and hydrologists[69]. Without this dialogue, there is a risk that the input used in hydrologists' models may contradict the meteorologists' assessments.

Traditionally, modelled hydrographs have been made presentable for users by manually adjusting model inputs, states, and parameters. This reduces the occurrence of implausible forecasts coming from atypical combinations of model states and forcings. Working within the model space often aids in preserving the physical consistency of all aspects of the forecasts, including other sites downstream. The research community has not adequately demonstrated practical methods of objective data assimilation that performs to forecaster expectations [3]. As a result, automatic data assimilation is very rare in operational flood forecasting, even though hydrological models are simpler and datasets are much smaller than their meteorological counterparts. The solution may lay with automatic data assimilation, manual post-adjustment of local features on the hydrograph, automated algorithms to maintain internal consistency, and automated checking that the forecaster is not over-adjusting.

Several countries are increasingly centralizing their regionalized river forecasting services into a national center. The broader geographic domain means that forecasters in a centralized office can develop experience more rapidly, as opposed to the forecasters in a small regional office, who may experience only a few extreme events in a career. The rarity of local learning opportunities is particularly challenging in dry climates. However, the increased operational workload in a centralized system may also have its drawbacks, including fewer opportunities to develop local knowledge or to interface with local customers. Because water management and other human impairments feature in most watersheds in many countries, and river forecasting relies on interaction with water managers, the effort to centralize forecasting operations must somehow leverage the information arising from local interactions. This information is most important for flood forecasting timescales (where water managers can control structures to alter the flow), and is relatively less important on seasonal and flash flooding timescales.

Interestingly, while very short and long leadtime forecasts are well suited for automation, there is also a rising trend towards the provision of "seamless" forecasts which cover all timescales (from hours to weeks) in a consistent manner. This could lead to automated products overlapping with semi-manual flood forecasts, and agencies may need to manage the communication issues that arise from potentially conflicting messages. This might be done by labelling the automated forecasts as experimental and providing links to the official (i.e. expert enhanced) forecasts.

5. Conclusions

Operational agencies are approaching a crossroad, where crucial decisions must be made about the role of hydrologists and automation in the production of river forecasts. There are many technology-driven opportunities to improve current operational river forecasts, such as access to better and more detailed weather and climate forecasts, improved connectivity, more complex models, and more powerful computers. The frontiers of scientific research are exploring methods such as data assimilation, statistical post-processing, and multi-model combination. Furthermore, agencies are facing increased sophistication and specialization of consumers and their requirements. Some of these advances do not fit well with traditional formalized practices of operational forecasting in hydrology.

It would be misguided to simply automate some of the hydrologists' existing tasks in a piecemeal fashion, without careful consideration of the consequences. The forecasting process may have to be redesigned to make the best use of the strengths of both, people and machines. The pitfalls of automation, such as the de-skilling of forecasters and the difficulty with reclaiming control when models fail, can be avoided through conscious design of the human-machine mix.

This article reviewed findings from the cognitive psychology and decision support systems literature, as well as results from other forecasting enterprises, such as meteorology. However, hydrology has several factors that differentiate it from other fields and the implications of such differences were also discussed. In particular, several domains of hydrology more amenable to automation such as flash flood forecasting and extended (sub-seasonal to seasonal) hydrologic prediction. Short-range flood forecasting, may require a more thoughtful and cautious pathway to automation. In this case, human interference in the hydrologic cycle and complex patterns of vulnerability mean that there is much "soft information" that must be considered for the forecasts to be effective.

Already, several automated forecasting systems are emerging and will continue to evolve. These systems should be encouraged to share their lessons with the hydrological community as part of a broader effort to engage forecasters, stakeholders and researchers in an ongoing conversation about balancing tradition and progress.

6. Acknowledgements

Thanks are given to Albert Mannes, Michael Cranston, Andrew Preece, Lionel Berthet, Jeff Perkins, Peter May, Robin Webb, Roger Deslandes, Bodo Zeschke, and Dasarath Jayasuriya for their reviews of this article and the Bureau technical report that led to the article's development. We are also appreciative of Tanya Smith's editorial improvements.

7. References

1. Cloke HL, Pappenberger F. Ensemble flood forecasting: a review. *Journal of Hydrology* 2009, 375:613-626.

2. Liu Y, Weerts A, Clark M, Hendricks Franssen H, Kumar S, Moradkhani H, Seo D, Schwanenberg D, Smith P, van Dijk A. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci. Discuss* 2012, 9:3415-3472.
3. Demargne J, Wu L, Regonda S, Brown J, Lee H, He M, Seo D-J, Hartman R, Herr HD, Fresch M. The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bulletin of the American Meteorological Society* 2013, 95:79-98.
4. Kahneman D, Klein G. Conditions for intuitive expertise: A failure to disagree. *American Psychologist* 2009, 64:515-526.
5. Nicholls N. Cognitive illusions, heuristics, and climate prediction. *Bulletin of the American Meteorological Society* 1999, 80:1385-1398.
6. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *International Journal of Human-Computer Studies* 1999, 51:991.
7. Bureau d'Enquetes et d'Analyses. Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro–Paris. BEA 2012. Available at: http://www.bea.aero/en/enquetes/flight.af.447/rapport.final.en.php.
8. IAEA. The role of automation and humans in nuclear power plants International Atomic Energy Agency 1997. Available at: http://www-pub.iaea.org/books/IAEABooks/930/The-Role-of-Automation-and-Humans-in-Nuclear-Power-Plants-Report-Prepared-Within-the-Framework-of-the-International-Working-Group-on-Nuclear-Power-Plant-Control-and-Instrumentation.
9. Doswell III CA. Weather forecasting by humans-Heuristics and decision making. *Weather and Forecasting* 2004, 19:1115-1126.
10. Sills DML. On the MSC forecasters forums and the future role of the human forecaster. *Bulletin of the American Meteorological Society* 2009, 90:619-627.
11. Novak DR, Bright DR, Brennan MJ. Operational forecaster uncertainty needs and future roles. *Weather and Forecasting* 2008, 23:1069-1084.
12. Stuart NA, Market PS, Telfeyan B, Lackmann GM, Carey K, Brooks HE, Nietfeld D, Motta BC, Reeves K. The Future of Humans in an Increasingly Automated Forecast Process. *Bulletin of the American Meteorological Society* 2006, 87:1497-1502.
13. Stuart NA, Schultz DM, Klein G. Maintaining the role of humans in the forecast process. *Bulletin of the American Meteorological Society* 2007, 88:1893-1898.
14. Sene K. *Hydrometeorology: forecasting and applications*. New York: Springer; 2010.
15. Fine GA. *Authors of the Storm: Meteorologists and the Culture of Prediction*. Chicago: University of Chicago Press; 2007.
16. Persson A, Grazzini F. User Guide to ECMWF forecast products. *Meteorological Bulletin* 2005, 3:153-154.
17. Pappenberger F, Cloke HL, Persson A, Demeritt D. HESS Opinions" On forecast (in) consistency in a hydro-meteorological chain: curse or blessing?". *Hydrology and Earth System Sciences* 2011, 15:2391-2400.
18. Pappenberger F, Stephens E, Thielen J, Salamon P, Demeritt D, Andel SJ, Wetterhall F, Alfieri L. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrological Processes* 2012, 27:132-146.
19. Pagano TC. Evaluation of Mekong River Commission operational flood forecasts, 2000-2012. *Hydrol. Earth Syst. Sci. Discuss.* 2013, 10:14433-14461.
20. Welles E, Sorooshian S, Carter G, Olsen B. Hydrologic Verification: A Call for Action and Collaboration. *Bulletin of the American Meteorological Society* 2007, 88:503-511.
21. Pielke Jr RA. Who decides? Forecasts and responsibilities in the 1997 Red River flood. *Applied Behavioral Science Review* 1999, 7:83-101.

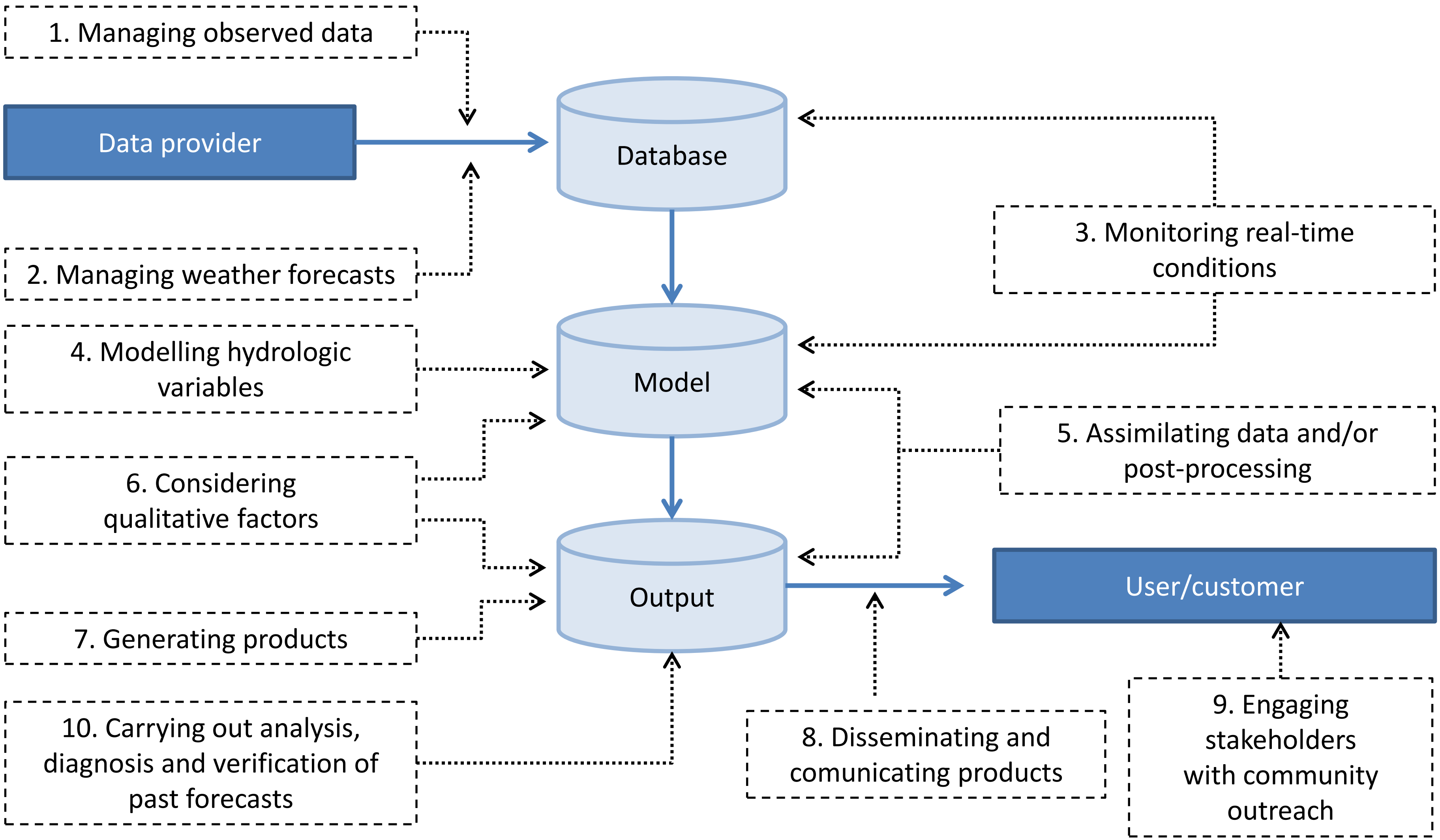
22. Pagano TC, Wood AW, Ramos M-H, Cloke HL, Pappenberger F, Clark MP, Cranston M, Kavetski D, Mathevet T, Sorooshian S, et al. Challenges of Operational River Forecasting. *Journal of Hydrometeorology* 2014, 15:1692–1707.
23. Pagano TC. International Review of the Role of Automation in River Forecasting Systems. Bureau of Meteorology 2013. Available at: https://www.researchgate.net/publication/255738177_International_Review_of_the_Role_of_Automation_in_River_Forecasting_Systems.
24. Practical Action. Early Warning Saving Lives. Available at: http://practicalaction.org/docs/region_nepal/early-warning-saving-lives.pdf.
25. Wood AW, Schaake JC. Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. *Journal of Hydrometeorology* 2008, 9:132-148.
26. Franz KJ, Hogue TS, Sorooshian S. Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. *Journal of Hydrology* 2008, 360:48-66.
27. Pagano TC, Wood AW, Werner K, Tama-Sweet R. Western U.S. Water Supply Forecasting: A Tradition Evolves. *Eos, Transactions American Geophysical Union* 2014, 95:28-29.
28. Pagano TC, Garen DC, Perkins TR, Pasteris PA. Daily updating of operational statistical seasonal water supply forecasts for the western U.S. *Journal of the American Water Resources Association* 2009, 45:767-778.
29. Adams T, Ostrowski J. Short lead-time hydrologic ensemble forecasts from numerical weather prediction model ensembles. In: *World Environmental and Water Resources Congress*. Providence, RI: American Society of Civil Engineers; 2010.
30. Price D, Hudson K, Boyce G, Schellekens J, Moore RJ, Clark P, Harrison T, Connolly E, Pilling C. Operational use of a grid-based model for flood forecasting. *Proceedings of the ICE-Water Management* 2012, 165:65-77.
31. Werner M, Cranston M, Harrison T, Whitfield D, Schellekens J. Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteorological Applications* 2009, 16:13-22.
32. Thielen J, Bartholmes J, Ramos M-H, De Roo A. The European Flood Alert System--Part 1: Concept and development. *Hydrology and Earth System Sciences* 2009, 13:125.
33. Alfieri L, Burek P, Dutra E, Krzeminski B, Muraro D, Thielen J, Pappenberger F. GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci. Discuss.* 2012, 9:12293-12332.
34. Bogner K, Pappenberger F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research* 2011, 47:W07524.
35. Hong Y, Adler RF, Hossain F, Curtis S, Huffman GJ. A first approach to global runoff simulation using satellite rainfall estimation. *Water Resources Research* 2007, 43.
36. de Winter J, Dodou D. Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work* 2011, 16:1-11.
37. Sheridan TB. Task analysis, task allocation and supervisory control. In: Helander M, ed. *Handbook of human-computer interaction*; 1997, 87-105.
38. Ayton P, Fischer I. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & cognition* 2004, 32:1369-1378.
39. Stewart TR. Improving reliability of judgmental forecasts. In: Armstrong JS, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. New York: Springer Science & Business Media; 2001, 81-106.
40. McKenzie CR, Liersch MJ, Yaniv I. Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes* 2008, 107:179-191.
41. Mannes AE, Moore DA. A Behavioral Demonstration of Overconfidence in Judgment. *Psychological science* 2013, 24:1190-1197.

42. Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* 1993, 8:281-293.
43. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 2000, 30:286.
44. Dawes RM. The robust beauty of improper linear models in decision making. *American Psychologist* 1979, 34:571-582.
45. Doswell III CA. Is there a role for humans in the NWS of the future? (Is there an NWS in our future at all?). Available at: http://webserv.chatsystems.com/~doswell/forecasting/human_role/future_forecasters.html.
46. Mosier KL, Palmer EA, Degani A. Electronic checklists: Implications for decision making. In: *Human Factors and Ergonomics Society Annual Meeting*. Atlanta, GA: SAGE Publications; 1992.
47. Parasuraman R, Riley V. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 1997, 39:230.
48. Muir BM, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 1996, 39:429-460.
49. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 2004, 46:50-80.
50. Pappenberger F, Ramos M-H, Cloke HL, Wetterhall F, Alfieri L, Bogner K, Mueller A, Salamon P. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology* 2015, 522:697-713.
51. Craven JP, Wiedenfeld J, Gagan J, Browning P, Just A, Greif C. The NWS Central Region Extended Forecast Process. In: *38th National Weather Association Annual Meeting*. Charleston, SC: National Weather Association; 2013.
52. Snellman LW. Operational forecasting using automated guidance. *Bulletin of the American Meteorological Society* 1977, 58:1036-1044.
53. Daipha P. Visual perception at work: Lessons from the world of meteorology. *Poetics* 2010, 38:151-165.
54. Ruth D. Interactive forecast preparation—the future has come. Preprints. In: *Interactive Symposium on the Advanced Weather Interactive Processing System (AWIPS)*. Orlando, FL: American Meteorological Society; 2002.
55. Doswell III CA, Maddox RA. The role of diagnosis in weather forecasting. In: *11th Conf. on Weather Forecasting and Analysis*. Kansas City, MO: American Meteorological Society; 1986.
56. Roebber PJ, Schultz DM, Colle BA, Stensrud DJ. Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Weather and Forecasting* 2004, 19:936-949.
57. McCarthy PJ, Ball D, Purcell W. Project Phoenix: Optimizing the machine–person mix in high-impact weather forecasting. In: *22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*. Park City, UT: American Meteorological Society; 2007.
58. Doswell III CA. The human element in weather forecasting. *Natl. Wea. Dig* 1986, 11:6-17.
59. Pagano TC, Garen D, Sorooshian S. Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922-2002. *Journal of Hydrometeorology* 2004, 5:896-909.
60. Hapuarachchi HAP, Wang QJ, Pagano TC. A review of advances in flash flood forecasting. *Hydrological Processes* 2011, 25:2771–2784.
61. Verkade J, Brown J, Reggiani P, Weerts A. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology* 2013:73-91.

62. Zalachori I, Ramos M, Garçon R, Mathevet T, Gailhard J. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science & Research* 2012, 8:135-141.
63. Schaake JC, Hamill TM, Buizza R, Clark M. HEPEX: The Hydrological Ensemble Prediction Experiment. *Bulletin of the American Meteorological Society* 2007, 88:1541-1547.
64. Pagano TC, Shrestha DL, Wang Q, Robertson D, Hapuarachchi P. Ensemble dressing for hydrological applications. *Hydrological Processes* 2012, 27:106-116.
65. Rossa A, Haase G, Keil C, Alberoni P, Ballard S, Bech J, Germann U, Pfeifer M, Salonen K. Propagation of uncertainty from observing systems into NWP: COST-731 Working Group 1. *Atmospheric Science Letters* 2010, 11:145-152.
66. Waser J, Ribicic H, Fuchs R, Hirsch C, Schindler B, Bloschl G, Groller M. Nodes on ropes: A comprehensive data and control flow for steering ensemble simulations. *Visualization and Computer Graphics, IEEE Transactions on* 2011, 17:1872-1881.
67. Blöschl G. Flood warning-on the value of local information. *International Journal of River Basin Management* 2008, 6:41-50.
68. Cuo L, Pagano TC, Wang Q. A Review of Quantitative Precipitation Forecasts and Their Use in Short-to Medium-Range Streamflow Forecasting. *Journal of Hydrometeorology* 2011, 12:713-728.
69. Ramos M-H, Mathevet T, Thielen J, Pappenberger F. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications* 2010, 17:223-235.

Related WIREs Articles

DOI	Article title
WATER-213.R1 (in press)	*Continental and Global Scale Flood Forecasting Systems*
10.1002/wcs.47	Expertise
10.1002/wat2.1075	Understanding the roles of modernity, science, and risk in shaping flood management
10.1002/wcc.187	Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction


1. Managing observed data
Data provider
Database
2. Managing weather forecasts
3. Monitoring real-time conditions
4. Modelling hydrologic variables
Model
5. Assimilating data and/or post-processing
6. Considering qualitative factors
Output
User/customer
7. Generating products
10. Carrying out analysis, diagnosis and verification of past forecasts
8. Disseminating and comunicating products
9. Engaging stakeholders with community outreach