Journal of Hydrology 551 (2017) 365–373

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

How does the quantification of uncertainties affect the quality and value of flood early warning systems?

A. Thiboult [a,*], F. Anctil [a], M.H. Ramos [b]

[a] Dept. of Civil and Water Engineering, Université Laval, Québec, Canada
[b] Irstea, UR HBAN, Antony, France

ARTICLE INFO

Article history:
Received 1 November 2016
Received in revised form 5 May 2017
Accepted 8 May 2017
Available online 15 May 2017
This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Helge Bormann, Associate Editor

Keywords:
Forecast quality
Relative economic value
Uncertainty estimation
Ensemble prediction
Data assimilation
Ensemble Kalman Filter
Multimodel

ABSTRACT

In an operational context, efficient decision-making is usually the ultimate objective of hydrometeorological forecasts. Because of the uncertainties that lay within the forecasting process, decisions are subject to uncertainty. A better quantification of uncertainties should provide better decisions, which often translate into optimal use and economic value of the forecasts. Six Early Warning Systems (EWS) based on contrasted forecasting systems are constructed to investigate how the quantification of uncertainties affects the quality of a decision. These systems differ by the location of the sources of uncertainty, and the total amount of uncertainty they take into account in the forecasting process. They are assessed with the Relative Economic Value (REV), which is a flexible measure to quantify the potential economic benefits of an EWS. The results show that all systems provide a gain over the case where no EWS is used. The most complex systems, i.e. those that consider more sources of uncertainty in the forecasting process, are those that showed the most reduced expected damages. Systems with better accuracy and reliability are generally the ones with higher REV, even though our analysis did not show a clear-cut relationship between overall forecast quality and REV in the context investigated.

1. Introduction

Floods are one of the most devastating natural disasters in the world (CRED and UNISDR, 2015). Related socio-economic impacts are considerable and require adequate prevention measures. Governments and communities seek to reduce the risk of floods, notably when the societal vulnerability is high in urban and industrial zones or when environmental and agricultural areas need to be protected, by locally implementing flood mitigation measures.

Traditionally, risk reduction is preferred over relief for economic and human considerations (Rogers and Tsirkunov, 2010). Skillful Early Warning Systems (EWS) have the capability to offer flood prevention by issuing warnings up to several days before the flood event. Recent studies have demonstrated that flood warnings are economically efficient (e.g. Priest et al., 2011; Molinari and Handmer, 2011; Verkade and Werner, 2011; Perrels et al., 2013). Frei (2010) estimated that benefits generated by weather services in Switzerland amount to some hundreds of millions of US$ per year. Similar results were obtained by Anaman and Lellyett (1996), Lazo and Chestnut (2002), Leviäkangas et al. (2007) in other industrialized countries, with ratios of invested and saved money ranging between 1:4 and 1:6. Pappenberger et al. (2015) evaluated that the European Flood Awareness System (EFAS, Thielen et al., 2009; Bartholmes et al., 2009), which provides information to national authorities and to the Emergency Response Coordination Center of the European Commission up to 15 days ahead, reaps benefits as high as 400 Euros for every 1 Euro invested.

In order to be valuable, forecasts from an EWS need to integrate the uncertainties inherent to the forecasting procedure. These uncertainties should reflect the inaccuracies that lay in the mathematical representation of the hydrometeorological system and in our knowledge of the initial and future states of the system (e.g. Ajami et al., 2007; Salamon and Feyen, 2010; Liu and Gupta, 2007; Liu et al., 2012). Ensemble prediction systems (EPS) provide a practical answer to incorporate different sources of uncertainty in the forecasting process. This approach has gained popularity and has been increasingly used by operational agencies (see the review by][and the online portal of the HEPEX community at www.hepex.org Cloke and Pappenberger, 2009). In a decision-making context, EPS-based forecasting systems have proved to be efficient and capable of improving the forecast value upon traditional deterministic forecasts (e.g. Richardson, 2000; Zhu et al.,

* Corresponding author.
E-mail address: antoine.thiboult.1@ulaval.ca (A. Thiboult).

http://dx.doi.org/10.1016/j.jhydrol.2017.05.014
0022-1694/© 2017 Elsevier B.V. All rights reserved.

2002; Verkade and Werner, 2011; Boucher et al., 2012; Stephens and Cloke, 2014), even if the communication of probabilistic forecasts in real-time remains a challenge (Ramos et al., 2010; Demeritt et al., 2013; Pagano et al., 2014).

Many authors have studied the quality of EPS-based forecasts, with first a focus on comparing deterministic and EPS forecasts. Forecast quality assessment has then evolved to assess the improvements in the quality of the forecasts when data-assimilation or post-processing techniques were applied to better quantify forecast uncertainty (e.g., Reggiani et al., 2009; Weerts et al., 2011; Bourgin et al., 2014; Boucher et al., 2015; Roulin and Vannitsem, 2015). Although studies assessing forecast quality are numerous, the assessment of the economic value of EPS-based flood forecast is still rare. In general, existing works have investigated how economic gains vary from using deterministic forecasts in a decision-making model that maximizes gains or minimizes losses over time, with respect to using probabilistic forecasts (Roulin, 2007; McCollor and Stull, 2008; Van den Bergh and Roulin, 2010; Muluye, 2011; Verkade and Werner, 2011; Boucher et al., 2012)). Additionally, the economic value of a forecast or a forecasting system is often tackled alone and the relation between quality and value is rarely addressed. As Verkade and Werner (2011) point out, it is expected that these aspects, quality and value, are linked, but more efforts should be put into clarifying what quality attributes of a forecast need to be improved to also improve its value.

Attempts to realistically define the damage associated to a particular river stage and the avoided losses from flood prevention measures are subject to many approximations and errors, and limited by the definition of the spatial and temporal boundaries of the event (Merz et al., 2010). The intangible costs resulting from deaths or traumas, for instance, are hard to quantify economically. When damages are tangible, the evaluation of costs is more straightforward but also subject to approximations since it may be difficult to take into account the indirect consequences of floods. To properly quantify damages, the approach to be adopted has to be sufficiently holistic to encompass all effective consequences, which can be social, political, and environmental (Merz et al., 2010).

To assess the economic gains related to protected values, Parker et al. (2007) use an estimation of the proportion of moveable inventory within a property. The main limitation of this approach is the fact that there are plenty of other measures, potentially more efficient to prevent damage losses. Moreover, flood warnings are not systematically followed by the population and efficient preventing measures are not always taken. More generally, decisions are made under constraints and can be encumbered by cues that are fallible, ambiguous, and altered by judgment (Choo, 2009).

The challenges mentioned above called for the use of the concept of maximum potential reduction of flood damage, which relates the actual flood damage avoided to other factors that stand in the way of optimal mitigation (Parker, 1991). The relation is defined as the product of the maximal potential reduction for a perfect system, the probability that the forecast is issued sufficiently in advance to react, the fraction of concerned people that will respond to the warning, and the fraction of people who will take effective measures. This product is estimated to 0.5 in the UK by the Department for Environment, Food and Rural Affairs (Verkade and Werner, 2011).

In order to avoid the cumbersome calculation of the maximum potential reduction damage, which would be particular to each studied location and would require a high number of approximations, the Relative Economic Value (*REV*) and the cost-loss ratio are often used. They are suitable to compare more easily different forecasting systems and to apply the methodology systematically to a large dataset of catchments. The *REV* is a more theoretical assessment of the value of a forecast. It is not based on real damage statistics but it can be easily transferable to more practical cases.

This study investigates how the quantification of uncertainties affects the quality of a decision. Six EWS were created using a framework hereafter named HOOPLA (HydrOlOgical Prediction Laboratory), which is a collection of hydrometeorological tools that allows constructing forecasting systems of various levels of complexity and sophistication. A simple framework is adopted to evaluate the economic gain that could be reached by the six EWS of different forecast quality. These systems differ by the way they take into account the main sources of uncertainty that play a role in hydrometeorological forecasting and, thus by the amount of total uncertainty they handle. As a result, they vary in terms of forecast performance, with different degrees of forecast accuracy, and reliability. We investigate their economic value and the contribution of their uncertainty components. The framework provides an estimate of the system's complexity required to take "better" decisions. From the results obtained, we also investigated if the quality of a hydrometeorological forecasting system, measured by typical scores, can be directly related to its economic value, as measured by the relative economic value (*REV*).

Section 2 presents the methodology, including the hydrometeorological data, the framework for the REV assessment, and the forecasting systems investigated. Results are presented in Section 3, where the REV and the relation between forecast quality and value are assessed. Concluding remarks are provided in Section 4.

2. Data and methodology

2.1. Catchment dataset and hydrometeorological data

The hydrometeorological dataset is composed of 20 catchments situated in the Province of Québec, Canada (Fig. 1). On these catchments, snow accumulation and melting are driving processes that create a spring freshet, while a second rain-induced flood peak may occur during fall. The catchment size and the mean annual streamflow vary from 512 to 15342 km^2 and from 8 to 300 $m^3{\cdot}s^{-1}$, respectively.

2.1.1. Cost-loss ratio

The cost-loss ratio (*CLR*) represents the ratio of the costs of mitigation and the avoidable losses due to an adverse event. It is defined as:

$$CLR = r = \frac{C}{L_a} \tag{1}$$

where C is the cost of the warning response, and L_a the avoidable losses. In the following, results are presented for values of *CLR* comprised within the interval $0 < CLR \leqslant 1$, as, economically, it does not make sense to take preventive measures that are more expensive than avoidable damages. Moreover, the *CLR* cannot equal 0 since operating an EWS already implies some costs. A hypothetical case of *CLR* equal zero would imply that the system could benefit from a continuous warning that has no cost. In practice, the cost-loss values are situated in a narrower range, but with the use of the aforementioned interval for the *CLR*, we can draw a more general conclusion based on our different EWSs and study areas. This range of CLR is also more convenient for our purposes as it can theoretically encapsulate different costs (e.g., costs to set/initialize the EWS, costs of operation, and costs associated with the mitigation of the event) and all sources of avoidable loss. Therefore, a wide range of potential cases can be built upon this synthetic assessment of the value of *CLR*.

2.1.2. Relative economic value

The Relative Economic Value (*REV*) is a dimensionless factor that scales between the case where no forecast is issued (thus no

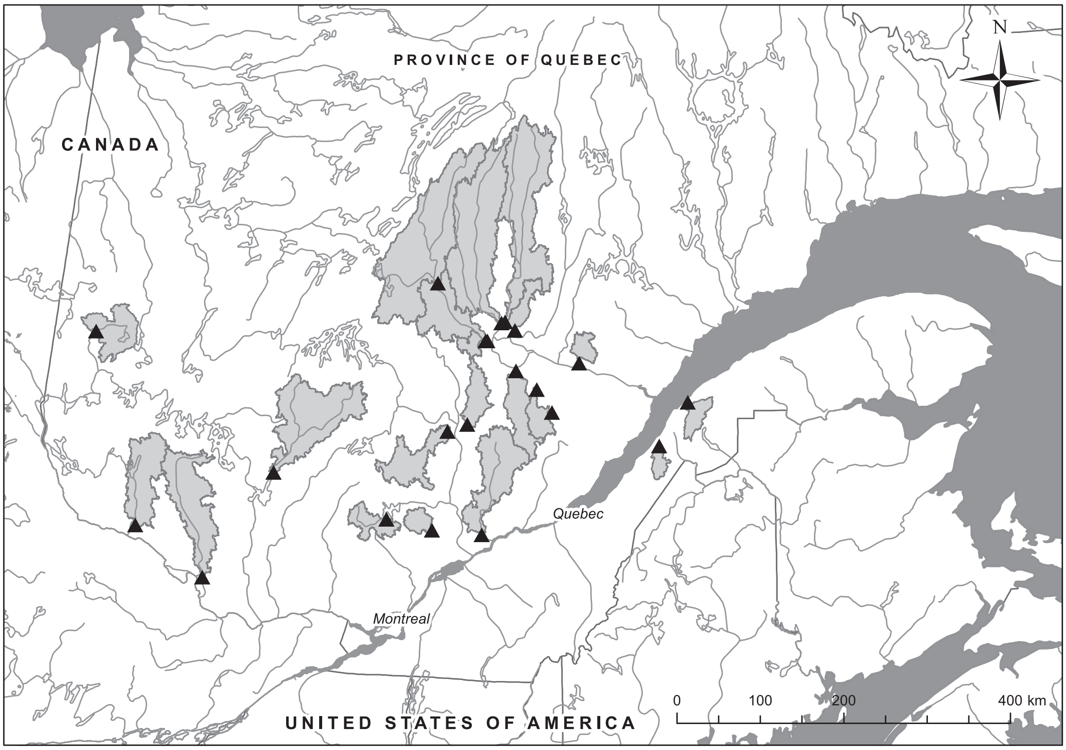

Fig. 1. Location of the 20 catchments.

warning) and the perfect forecast case where a warning is issued for every adverse event (Zhu et al., 2002). A *REV* equals to 1 denotes the best possible forecast decision system, while a *REV* equals to 0 indicates that the system is equivalent to the no forecast case. In other words, the *REV* is positive whenever the use of the corresponding EWS is beneficial and negative otherwise.

The costs associated with an EWS can be estimated from a contingency table. Table 1 describes the four possible cases and their frequency of occurrence, hits (h), false alarms (f), missed events (m) and correct negatives (c), and the different types of costs and losses associated with each outcome, i.e. avoidable loss (L_a), unavoidable loss (L_u), and warning cost (C).

In the no warning case, the expected costs over the assessment period are given by the consequences of a missed event, i.e. the sum of avoidable loss, multiplied by the number of times an event was observed $h+m$:

$$E_{nowarn} = (h+m)(L_a + L_u) \tag{2}$$

The no forecast case can also be seen as a system that has the same skill as the climatology. Since the frequency of river stages exceeding the critical threshold is statistically low, the user will not issue a warning if no forecast information is available (Verkade and Werner, 2011).

Table 1
Contingency table with costs associated with each type of event.

		Warning issued	
		Yes	No
Event observed	Yes	Hit (h) Mitigated Loss ($C+L_u$)	Miss (m) Loss (L_a+L_u)
	No	False Alarm (f) Warning Cost (C)	Correct negative (c) No Cost (–)

A perfect forecasting system will always give warnings when events are observed. There will be no misses or false alerts. The expected damages are thus the product of the event occurrences $(h+m)$ and the mitigated losses associated with the hit outcomes:

$$E_{perfect} = (h+m)(C+L_u) \tag{3}$$

Finally, in general, the expected costs for an early warning system will include all possible outcomes and their consequences:

$$E_{EWS} = h(C+L_u) + m(L_a+L_u) + fC \tag{4}$$

The expected costs obtained with the three cases (i.e., the EWS, the no warning, and the perfect warning case) can be put together using the *REV* framework. The dimensionless *REV* score is the equivalent of a skill score that scales between the optimal value $E_{perfect}$ and the reference value E_{nowarn}.

$$\begin{aligned} REV &= \frac{E_{nowarn} - E_{EWS}}{E_{nowarn} - E_{perfect}} = \frac{(h+m)L_a - (h+f)C - m(L_a)}{(h+m)L_a - (h+m)C} \\ &= \frac{(h+m) - (h+f)r - m}{(h+m)(1-r)} \end{aligned} \tag{5}$$

2.1.3. Optimal decision rule

The above decision-making framework can be applied to deterministic and probabilistic hydrological forecasting systems. If the decision to issue a warning is solely based on mathematical evidence (i.e., human expertise of model outputs are not considered), the forecaster will issue a warning if the expected damages with a mitigation action are lower than the expected losses without mitigation. For a probabilistic forecast, the optimal decision rule is a function of the probability P of exceeding a given flood threshold (above which flooding occurs):

$$\begin{aligned} C + PL_u &< P(L_a + L_u) \\ \frac{C}{L_a} &< P \\ r &< P \end{aligned} \tag{6}$$

For a deterministic forecast, the probability of exceeding the threshold is either 0, if the forecast is under a given flood threshold, or 1, if it is above the threshold. For an EPS-based forecast, it is estimated by the number of members above the threshold over the total number of members of the ensemble.

2.2. Flood threshold

In practical cases, flood thresholds are expressed in terms of river stages. However, as we do not possess rating curves nor critical river stages, we define a streamflow threshold and make the common assumption that at each stage corresponds a unique streamflow.

A threshold is set for each catchment and defined as the quantile 0.9 of the observed streamflows on the assessment period (i.e., 10% of the streamflow values of the assessment period are greater than the threshold). We thus have the same number of adverse events on all catchments.

2.3. Scores of forecast performance

The accuracy of the forecasts is measured by the Mean Absolute Error (*MAE*) computed on the mean of the ensemble. A value of 0 corresponds to a perfect match of the observed streamflow and the mean of the forecast ensemble.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^{obs}| \tag{7}$$

where y_i is the streamflow forecast for day i and y_i^{obs} the corresponding observation.

Forecast reliability is assessed with the rank histogram. The rank histogram, also named the Talagrand histogram, is a graphical representation that illustrates the empirical frequencies of the positions of the observations within the members of the ensemble forecast, when all values are ranked. In other words, $d+1$ bins are delineated by the ranked values of the d members of the ensemble plus the observation. In the case of a perfectly reliable forecast, the N observations should fall as often in each bin, producing a flat histogram (or uniform distribution). The deviation from the histogram flatness can be measured by:

$$\Delta = \sum_{c=1}^{d+1}(S_c - h_{ref})^2 \tag{8}$$

where S_c is the number of observations in the bin c, and $h_{ref} = N/(d+1)$, the expected value. Thus, a perfectly reliable system is expected to have: $\Delta_{perfect} = (dN)/(d+1)$. Candille and Talagrand (2005) suggested the δ ratio to measure the reliability of a forecasting ensemble system from the ranked histogram:

$$\delta = \Delta/\Delta_{perfect} \tag{9}$$

A δ ratio equal 1 indicates a perfectly reliable forecast, while values much larger indicate a deviation from the perfectly reliable forecast.

2.4. Forecasting systems

Six EWSs that differ by the way uncertainties are quantified in their hydrometeorological forecasting system are investigated. They were created with the help of the HOOPLA framework, which allows to produce, calibrate and run forecasting systems of different levels of complexity. In order to build the systems, we considered a meteorological ensemble prediction system, a multi (hydrological) model approach, and the Ensemble Kalman Filter (EnKF) to initialize the states of the hydrological model. These options are hereafter referred as the (hydrometeorological) tools. The combination of the three tools leads to the six systems (Table 2). By using either none, one, or several of the tools, these systems are expected to decipher a different amount of the total predictive uncertainty.

Table 2
Description of the six systems.

Systems	A	B	C	D	E	F
Multimodel	Off	Off	Off	On	Off	On
EnKF	Off	Off	On	Off	On	On
Met. ensemble	Off	On	Off	Off	On	On
Nb of members	(20×)1	(20×)50	(20×)50	20	(20×)2500	50,000

The meteorological ensemble prediction system comes from the 50 perturbed members of the EPS of the European Center for Medium-Range Weather Forecasts (ECMWF), and was retrieved from the TIGGE database. We used precipitation and temperature over a 9-day forecast horizon and a 2 year-period (from November 2008 to December 2010). The predictions were aggregated temporally at a daily time step and spatially at the catchment scale. In systems B, E and F (Table 2), the 50 equiprobable members are processed through the hydrological models, otherwise a single member is randomly selected to represent a deterministic meteorological forecast.

The multimodel approach allows to estimate structural uncertainty in a dynamic way by considering the responses of different models. It is an attractive approach as it does not rely on past errors of a single model and therefore avoids the typical constraint of model error stationarity. Other approaches to quantify hydrological modelling uncertainty exist, based mainly on post-processing techniques (see, for instance, Bennett et al., 2015; Li et al., 2016; Roulin and Vannitsem, 2015). Regardless of the approach used, the quantification of hydrological uncertainty remains an imperfect approximation as it is practically impossible to fully sample all possible structural errors and representations which are likely to describe all hydrological processes in a catchment. The multimodel ensemble is a pool of twenty lumped, conceptual hydrological models. This multimodel ensemble was first studied by Perrin (2000) and revised by Seiller et al. (2012). The selection is based on the differences in the structures of the models and their degree of complexity, allowing to cover a broad range of possible conceptual descriptions of the hydrological processes. Some models were modified to match the needs of the multimodel frame and may differ from their original version. These modifications include, for instance, the spatial representation of the catchments or the model time step. The main structural specificities of each model were however preserved. The models are coupled with a 2-parameter degree-day snow accounting routine (Valery et al., 2014) and a potential evapotranspiration formula (Oudin et al., 2005). Table 3 shows the twenty models selected to be used in this study. Details about model each structure and performance can be found in Perrin (2000) and Thiboult and Anctil (2015).

In this study, the models were individually calibrated on a 10-year period using the *RMSE* on square-rooted streamflows as objective function. They were validated on a different 10-year sequence. Fig. 2 shows the performance of the models in validation with the Nash Sutcliffe Efficiency score (*NSE*, Nash and Sutcliffe, 1970). Model performance varies among the catchments (range in the boxplots), but none of the models outperforms the others in all catchments.

The Ensemble Kalman Filter (EnKF) is used on each model to reinitialize the model states prior to the forecast. The filter is tuned per catchment and per model for forecasting systems C and E. A lot of effort has been dedicated to identify the level of required

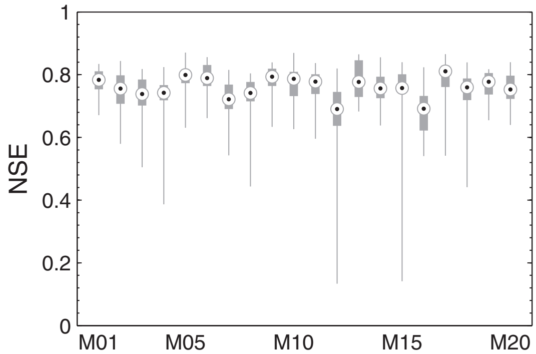

Fig. 2. Nash Sutcliffe Efficiency (*NSE*) values of the performance of the hydrological models (M1 to M20) in validation for all the 20 studied catchments (boxplot).

Table 3
Main characteristics of the 20 lumped models (Seiller et al., 2012)

Model acronym	No. of optimized. parameters[a]	Number of reservoirs	Derived from
M01	6	3	BUCKET (Thornthwaite and Mather, 1955)
M02	9	2	CEQUEAU (Girard et al., 1972)
M03	6	3	CREC (Cormary and Guilbot, 1973)
M04	6	3	GARDENIA (Thiery, 1982)
M05	4	2	GR4J (Perrin et al., 2003)
M06	9	3	HBV (Bergström and Forsman, 1973)
M07	6	5	HYMOD (Wagener et al., 2001)
M08	7	3	IHACRES (Jakeman et al., 1990)
M09	7	4	MARTINE (Mazenc et al., 1984)
M10	7	2	MOHYSE (Fortin and Turcotte, 2007)
M11	6	4	MORDOR (Garçon, 1999)
M12	10	7	NAM (Nielsen and Hansen, 1973)
M13	8	4	DFM (Moore and Clarke, 1981)
M14	9	5	SACRAMENTO (Burnash et al., 1973)
M15	8	3	SIMHYD (Chiew et al., 2002)
M16	8	3	SMAR (O'Connell et al., 1970)
M17	7	4	TANK (Sugawara, 1979)
M18	7	3	TOPMODEL (Beven et al., 1984)
M19	8	3	WAGENINGEN (Warmerdam et al., 1997)
M20	8	4	XINANJIANG (Zhao et al., 1980)

[a] To each model, one should add the 2 parameters of the snow routine.

perturbations to add to the meteorological inputs and hydrological observations, as well as to select the state variable that should be updated in order to maximize accuracy and reliability (Thiboult and Anctil, 2015). When the EnKF is combined with the multimodel approach (System F), it is tuned in a different way. The perturbations added to the inputs and outputs are the same for all models and catchments. If the EnKF is not used (i.e. no data assimilation is performed), it is referred as "open loop".

Table 2 summarizes the specificities of the six systems. System A, which is fully deterministic, should be seen as a benchmark for assessing the improvements brought by the use of the probabilistic tools to quantify uncertainty sources. Systems B, C and D make use of a single tool, while systems E and F combine two and three tools, respectively.

Note that systems A, B, C, and E do not rely on the hydrological multimodel approach but on a single structure instead. Consequently, one could generate 20 subsystems for each of the aforementioned systems: A(M01), A(M02), ..., A(M20), which include the hydrological models M01, M02, ..., M20, respectively. To isolate the improvements brought by the meteorological ensemble forcing and by the EnKF (systems B and C) over the benchmark (System A), a single model needs to be selected. For this purpose, we chose the model M14. It is the model that shows median performance (based on *NSE* values) during calibration and validation over all catchments.

3. Results

3.1. Relative economic value

Fig. 3 illustrates the *REV* for all systems and lead times of 1, 3, 5 and 9 days. Each color corresponds to a different system. The *REV* values are computed for a range of cost-loss ratios that varies between 0 and 1 with a 0.01 increment. Since the behaviour of the curves were very similar for all catchments, and no correlation was observed between the results and the main characteristics of the catchments, the time series of all 20 catchments were concatenated for the calculation of the contingency tables to enhance readability of the figures.

Fig. 3 shows that in the vast majority of cases the *REV* of every system decreases with increasing cost-loss ratio (*CLR*). This behavior is expected since the highest cost-loss ratios correspond to the most complex decision situations, where false alarms are increasingly expensive. Despite this trend, all systems are economically efficient (i.e., *REV* values are positive) for most *CLR* values. It is economically attractive to use any of these EWSs for all lead times, regardless of their complexity, when the cost-loss ratio is lower than 0.6.

The second finding is that *REV* values are not equal over all systems. System F shows the highest economic value for all lead times and *CLR*, and always exhibits positive *REV*, even for cost-loss ratios close to 1. By contrast, systems A and B offer the lesser improvement over the no-warning expected damage (i.e., they more often result in the lowest *REV* among the systems). Note that systems A and B, which only differ by the type of meteorological forecasts used (i.e., deterministic meteorological prediction for System A and ensemble prediction for System B), are not distinguishable for day 1. This means that picking randomly one meteorological member or using all members of the ensemble does not result in different economic values for the first day of lead time in the studied catchments. This can also be an indication that the spread of the ensemble forecasts is not large enough at day 1 to distinguish these two systems. Finally, System D is constantly ranked in the middle of the other systems.

With increasing lead times, the value of all EWS logically decreases as the number of false alarms and missed events increases. However, this happens at a different decreasing rate according to the system. The superiority of System F is confirmed as it exhibits higher *REV* than other systems for all lead times. System C, which uses only data assimilation, is often the second best one for short lead times, but experiences a substantial decrease and henceforth is outperformed by System D from day 5 onwards. System B, which was indistinguishable from System A for day 1, gains in relative efficiency for longer lead times and outperforms System A from days 4–5. This behavior is confirmed by the analysis of the meteorological ensemble (not shown here) where the dispersion is almost non-existent for day 1 and low for the following days. In summary, for the six systems investigated, meteorological ensemble forcing becomes valuable mostly for lead times greater than 4–5 days and the effects of data assimilation are stronger at shorter lead times.

The *REV* analysis indicates that the economic value of a system can be improved by accounting explicitly for the principal hydrometeorological sources of uncertainty, as it is usually observed in studies on forecast quality and EWS performance. The different

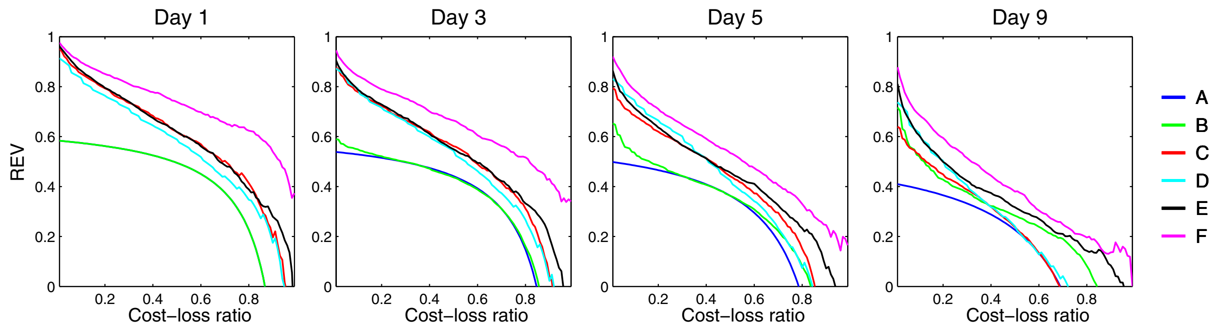

Fig. 3. Relative Economic Value (*REV*) according to the cost-loss ratio for the six early warning systems and days 1, 3, 5, and 9 of lead time. Note that values for systems A and B overlap for day 1. Results for 20 catchments and the forecasting period from 2008 to 2010.

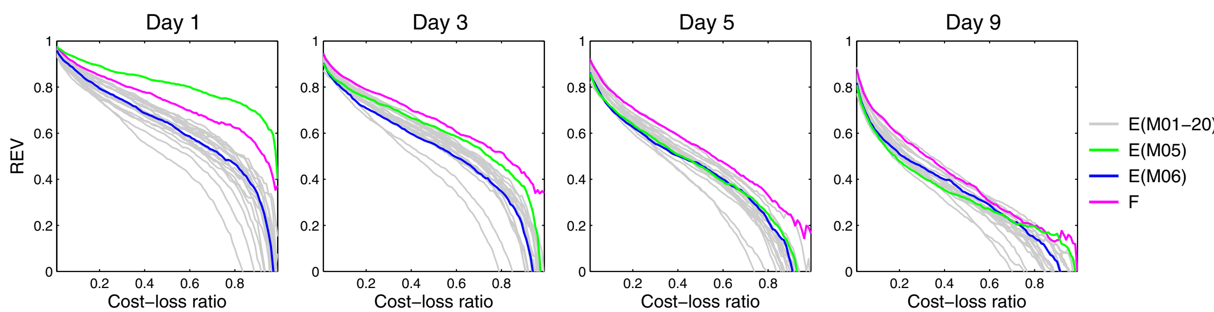

Fig. 4. Relative Economic Value (*REV*) in forecasting according to the cost-loss ratio of all models composing System E (M01 to M20) and System F for days 1, 3, 5, and 9. The two best performing models in validation (M05 and M06) are highlighted.

tools improve the *REV* for different lead times: data assimilation acts on shorter lead times since the influence of the EnKF eventually vanishes and may, in some cases, become worse than the open loop; the quantification of uncertainty from the use of meteorological ensemble forcing is attractive mostly from days 4–5 due to underdispersion at shorter lead times and the contribution of the multimodel appears to be quite constant over lead times. More complex systems, which take into account more sources of uncertainty by merging several tools, are overall more economically attractive, although they may also be more demanding in terms of development efforts, calibration, and computational requirements. One must also note that the differences in costs among the six systems are possibly small compared to the costs engaged in mitigation actions, avoidable damages or damage flood relief.

3.2. An attempt to simplify the most valuable systems

This section focuses on investigating further the role of the multimodel approach and the possible benefits from the selection of a single "best" hydrological model. While Section 3.1 presented the results of the median-performing model in validation for systems A, B, C, and E (model M14), this section investigates the *REV* of all individual models (models M01 to M20) of System E. System E was chosen as it was found to be the second-best system after System F (Section 3.1) and and it relies on tools that are more frequently used and available (EnKF and ensemble meteorological forcing). Two subsystems of System E are examined in details, M05 and M06, as these two models are the two best-performing ones in the validation process (Fig. 2).

Fig. 4 confirms the added value of the EWSs over the no-warning case for most cost-loss ratios even if, as expected, differences are observed in the individual REV of the 20 models (gray lines). The 20 models have performance values relatively close to each other in validation when used in open loop (Fig. 2), but System E(M05) clearly stands out in terms of *REV* in forecasting for the first lead time, when it even outperforms the more complex System F. However, the *REV* of System E(M05) decreases substantially with lead time. System E(M06), which has a *MAE* score in validation only 0.01 lower than System E(M05), exhibits a lower *REV* in comparison to the other models. System F remains the most valuable forecasting system in the vast majority of cases.

These results indicate that it is hazardous to rely on the validation performance of a hydrological model to maximize the expectations one can have in terms of its value when used in a forecasting system. Measures of the accuracy of a hydrological model in validation are a poor indicator of the economic value of a flood forecasting system based on this model. It is hence complex to identify a single hydrological model structure when its selection is performed "a priori" (i.e. in calibration or validation) and do not target the situations of interest when using the model for hydrological forecasting (here, flood forecasting).

It may be attractive to rely on a single hydrological model structure since it does not require multiple calibrations, such as for System F, and the use of a unique model is computationally lighter and easier to comprehend. However, such selection is risky when setting up a forecasting system. System F has drawbacks related to its higher complexity but it offers more consistant performance over lead times and catchments which is not equaled by any single hydrological structure. This consistancy is attributed to the way System F takes structural uncertainty into account. Structural uncertainty is solely accounted implicitly by data assimilation through state updating in System E, while it is accounted explicitly by the means of the multimodel in System F.

3.3. Relationship between value and quality

This section investigates the relationship between the *REV* of a forecasting system and its quality. Quality is measured by its

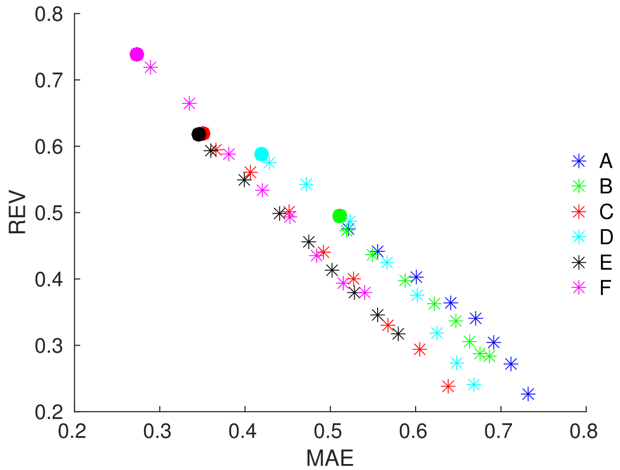

Fig. 5. Relationship between Mean Absolute Error (*MAE*) and the Relative Economic Value (*REV*) computed for a $CLR = 0.5$. The circles indicate day 1 results while asterisks indicate subsequent days.

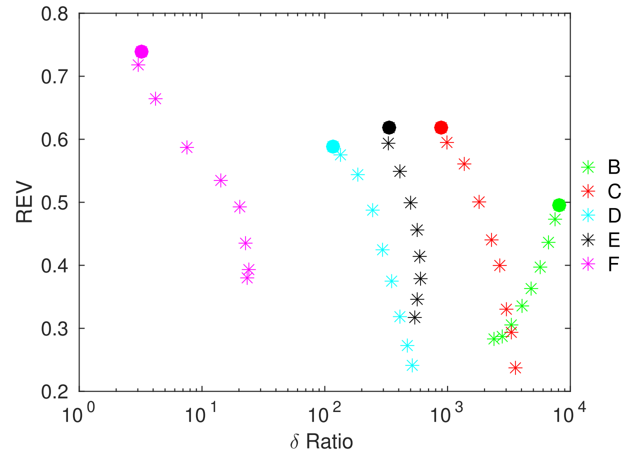

Fig. 6. Relationship between the δ ratio and the Relative Economic Value (*REV*) calculated for a $CLR = 0.5$. The circles indicate day 1 results while asterisks indicate subsequent days.

accuracy (*MAE*) and reliability (δ ratio). Other scores have been also tested for the computation of accuracy and reliability, without any change in the conclusions (for instance, we evaluated the continuous ranked probability score, the root mean squared error, the Nash–Sutcliffe efficiency calculated on both streamflow and square rooted streamflows for assessing accuracy, and several measures of deviation from perfect reliability in reliability plots). Here, values of *REV* are computed for a cost loss ratio equal to 0.5, which is a standard value, less subject to sampling errors (Richardson, 2001).

Fig. 5 shows the *REV* according to the *MAE*. All lead times are represented. Day-1 values are illustrated by a circle while the other forecast horizons (day 2 to 9) are denoted by asterisks. The decrease of the accuracy is monotonic (first lead time always has better accuracy than the second one, the second one better than the third one, etc), thus lead times are easily identifiable.

There is a clear correlation between *MAE* and *REV*. An improvement in the *MAE* value generally goes hand in hand with an improvement of the *REV*, indicating that the different systems behave coherently as more quality translates into more value. Therefore, the measure of forecast accuracy may be seen as an indicator of the economic value of a forecast. Nonetheless, when considering all systems together, similar *REVs* can be obtained for dissimilar *MAE* values. The relationship is better described by an envelope of possible values than by a single linear regression. These observations are in accordance with the findings from Murphy and Ehrendorfer (1987) and Richardson (2001), who performed a similar analysis in a theoretical framework with the Brier score.

Fig. 6 illustrates the distribution of the *REV* as a function of the δ ratio. System A is not represented as it is a deterministic system. The relationship between reliability and *REV* is not as striking as the relationship between accuracy and *REV*. Fig. 6 shows that it strongly depends on the EWS. The relationship can be positively oriented (better reliability corresponds to higher *REV*) like for Systems C and D, negatively oriented as it is the case for System B, or mostly positively oriented but with a more complex behavior, as for Systems E and F.

System B does not follow the same logic as the others, since its spread is only gradually generated by the meteorological forcing as lead time increases. The meteorological ensemble spread is almost zero for the first day and grows with lead time. This is reflected in the hydrological forecast ensemble generated, which also considerably lacks spread for the shorter lead times. As a consequence, its reliability improves considerably with increasing lead time, which explains the opposite orientation of the curve describing System B in Fig. 6. Note that System E also presents a similar behavior from day 6. At this forecast horizon, the spread generated by the EnKF data assimilation procedure vanishes and it is the spread of the ensemble meteorological forecasts that prevails. The decrease of the EnKF spread can be explained by the resilient nature of the hydrological models. When the perturbations to the states of the models are small (which is mostly the case during the data assimilation procedure), these states tend to converge towards the unperturbed values after a few iterations. This aspect is further discussed in Thiboult and Anctil (2015) and (Thiboult et al., 2016).

These observations indicate that the relationship between reliability and *REV* is not defined intrinsically as it depends on the forecasting system. More precisely, this relationship is defined by the way the system takes into account the different uncertainty sources of the total predictive uncertainty by using a specific combination of tools. In our study, a score that measures forecast reliability proved to be a good indicator of the capacity of a system to describe the total hydrometeorological uncertainty (the more uncertainty sources are quantified, the better the system), but it is a poor indicator of the economic value of a forecast for early warning (the same *REV* can be associated to a wide range of forecast reliability scores values).

It is also worth noticing that systems that have a better reliability are also, in general, the most accurate and valuable. This emphasizes that, at least with the tested systems, the different aspects of forecast quality are intrinsically related and cannot be fully separated (i.e. one cannot create a forecasting system that improves the scores of forecast reliability without changing the score values of forecast accuracy). The way the forecasting systems were developed in this study makes that an improvement in the description of forecast uncertainty with the use of the chosen hydrometeorological tools improves reliability and accuracy together (see also Thiboult et al., 2016; Thiboult et al., 2016 and Crochemore et al., 2016; Crochemore et al., 2016 for other examples of trade-offs between forecast quality attributs in streamflow forecasting). If a user is focused on forecast value under similar conditions as we have tested here, attention should be paid to accuracy, even if reliability also generally contributes to the *REV* of ensemble predictions.

4. Conclusion

This paper presents (i) a comparison of six Early Warning Systems (EWS) in terms of economic value, (ii) an analysis of the contribution of the multimodel approach, the data assimilation with Ensemble Kalman Filter, and the ensemble meteorological forcing to the economic value of EWS, and (iii) a preliminary analysis of the impact of improved accuracy and reliability on the economic value of a foreacast system. Each system investigated includes a specific set of forecasting components that differs from the others by the way it deciphers one or several sources of uncertainty and, therefore, by the amount of total hydrometeorological uncertainty it takes into account. The systems vary in complexity, ranging from a system based on a deterministic forcing and a single hydrological model to a system that includes multiple hydrological models, data assimilation with an ensemble Kalman filter and meteorological ensemble forcing.

The assessment of the forecast economic value relies on a flexible theoretical framework, where the relative economic value (*REV*) is used to scale the forecast value between the no-warning case and the perfect forecast case. Warnings are issued and subsequent costly mitigation actions are carried out when the forecast streamflow exceeds a predefined threshold.

Results show that all systems provide positive *REV* up to 9 days ahead for most cost-loss ratios. More complex systems provide higher economic value. By addressing specifically and adequately the three major sources of uncertainty in hydrometeorological modeling (i.e. initial condition, meteorological forcing, and structural uncertainty), the forecast economic value is increased for all lead times. The attempt to select a single hydrological model instead of using a multimodel approach did not result in higher economic values. This selection also proved to be complex, since the identification of the hydrological model to be used was not dependent on the validation performance criteria solely. In addition, a model that performs better for a given forecast lead time is not systematically the best for all lead times considered, which makes the choice of a "best" model structure not always unique in a forecasting context.

The preliminary investigation of a relationship between the quality and the value of a forecasting system revealed that, in general, better accuracy and reliability translates into higher economic values as measured by the *REV*. However, while the link was more clearly defined for the forecast accuracy attribute, the same was not observed for reliability. Our results showed that the relationship between the *REV* and the δ ratio strongly depends on the system and on how reliability and ensemble spread vary with lead time.

From our results, it was clear that a better quantification of forecat uncertainty increases the economic value of a forecat system. How improvements in forecast quality translate into higher economic values is however a topic that deserves further investigation. When improving a forecasting system, several attributes of forecast quality are simultaneously affected and it may be difficult to detect which attribute contributes the most to improve the value of a forecast. The search for a straightforward relationship between quality and value in a forecasting system may finally depend on the way the forecasts are going to be used in a decision-making problem. Our study focused on early warning systems for flood forecasting. Other forecast users may have different interests in terms of forecast quality and value, and a broader range of applications may provide more insights into the quality-value relationship in probabilistic streamflow forecasting. The six tested forecasting systems relied on a lumped and daily description of hydrological processes. Further work could be done to investigate also the impacts of uncertainty quantification on quality and value of flood forecasts at finer spatial and temporal resolutions. The latter aspect may be particularly important in the case of faster responding catchments and catchments prone to flash floods. Further work is also required to better take into account suitable mitigation measures, to identify actual flood damage avoided, and to investigate the possibility to improve decision rules in flood forecasting applications.

Acknowledgements

The authors acknowledge the Centre d'Expertise Hydrique du Québec for providing hydrometeorological data and the ECMWF for the development and maintenance of the TIGGE data portal. The two first authors also acknowledge financial support from the Chaire de recherche EDS en prévisions et actions hydrologiques and from the Natural Sciences and Engineering Research Council of Canada. The third author acknowledges the IMPREX research project, supported by the European Commission under the Horizon 2020 Framework programme, with grant nr 641811 (www.imprex.eu).

References

Ajami, N.K., Duan, Q.Y., Sorooshian, S., 2007. An integrated hydrologic bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43, 1–19.
Anaman, K.A., Lellyett, S.C., 1996. Assessment of the benefits of an enhanced weather information service for the cotton industry in australia. Meteorolog. Appl. 3, 127–135.
Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The european flood alert system efas - part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci. 13, 141–153.
Bennett, J.C., Wang, Q.J., Robertson, D.E., Schepen, A., 2015. Reliable long-range ensemble streamflow forecast: combining calibrated climate forecasts with a conceptual runoff model and a staged error model. Water Resour. Res. 52, 8238–8259.
Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff model. Nord. Hydrol. 4, 147–170.
Beven, K.J., Kirkby, M.J., Schofield, N., Tagg, A.F., 1984. Testing a physically-based flood forecasting model (topmodel) for 3 uk catchments. J. Hydrol. 69, 119–143.
Boucher, M.A., Perreault, L., Anctil, F., Favre, A.C., 2015. Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. Hydrol. Process. 29, 1141–1155.
Boucher, M.A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012. Hydro-economic assessment of hydrological forecasting systems. J. Hydrol. 416, 133–144.
Bourgin, F., Ramos, M.H., Thirel, G., Andreassian, V., 2014. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. J. Hydrol. 519, 2775–2784.
Burnash, R.J.C., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system – Conceptual modelling for digital computers.
Candille, G., Talagrand, O., 2005. Evaluation of probabilistic prediction systems for a scalar variable. Q. J. Royal Meteorolog. Soc. 131, 2131–2150.
Chiew, F.H.S., Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. pp. 335–367.
Choo, C.W., 2009. Information use and early warning effectiveness: perspectives and prospects. J. Am. Soc. Inform. Sci. Technol. 60, 1071–1082.
Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. J. Hydrol. 375, 613–626.
Cormary, Y., Guilbot, A., 1973. Étude des relations pluie-débit sur trois bassins versants d'investigation. In: publications, I. (Ed.), Design of water resources projects with inadequate data: IAHS Madrid Symposium. pp. 265–279.
CRED, UNISDR, 2015. The human cost of weather related disasters. Report.
Crochemore, L., Ramos, M., Pappenberger, F., 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrol. Earth Syst. Sci. 20, 3601–3618.
Demeritt, D., Nobert, S., Cloke, H.L., Pappenberger, F., 2013. The european flood alert system and the communication, perception, and use of ensemble predictions for operational flood risk management. Hydrol. Process. 27, 147–157.
Fortin, V., Turcotte, R., 2007. Le modèle hydrologique mohyse, note de cours pour sca7420. Report, Département des sciences de la terre et de l'atmosphère, Université du Québec à Montreal.
Frei, T., 2010. Economic and social benefits of meteorology and climatology in switzerland. Meteorolog. Appl. 17, 39–44.
Garçon, R., 1999. Modèle global pluie-débit pour la prévision et la prédétermination des crues. La Houille Blanche 7 (8), 88–95.
Girard, G., Morin, G., Charbonneau, R., 1972. Modèle précipitations-débits à discrétisation spatiale. Cahiers ORSTOM, Série Hydrologie 9, 35–52.
Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. J. Hydrol. 117, 275–300.
Lazo, J.K., Chestnut, L.G., 2002. Economic value of current and improved weather forecast in the u.s. household sector. Report.
Leviäkangas, P., Hautala, R., Räsänen, J., Öörni, R., Sonninen, S., Hekkanen, M., Ohlström, M., Venäläinen, A., Saku, S., 2007. Benefits of meteorological services in croatia. Finnish Meterological Institute, Helsinki. Report.

Li, M., Wang, Q.J., Bennett, J.C., Robertson, D.E., 2016. Error reduction and representation in stages (erris) in hydrological modelling for ensemble streamflow forecasting. Hydrol. Earth Syst. Sci. 20, 3561–3579.
Liu, Y., Weerts, A.H., Clark, M., Franssen, H.J.H., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol. Earth Syst. Sci. 16, 3863–3887.
Liu, Y.Q., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour. Res. 43, 1–18.
Mazenc, B., Sanchez, M., Thiery, D., 1984. Analyse de la influence de la physiographie d'un bassin versant sur les paramètres d'un modèle hydrologique global et sur les débits caractéristiques à la exutoire. J. Hydrol. 69, 97–188.
McCollor, D., Stull, R., 2008. Hydrometeorological short-range ensemble forecasts in complex terrain. part ii: Economic evaluation. Weather Forecasting 23, 557–574.
Merz, B., Kreibich, H., Schwarze, R., Thieken, A., 2010. Review article 'assessment of economic flood damage'. Nat. Hazards Earth Syst. Sci. 10, 1697–1724.
Molinari, D., Handmer, J., 2011. A behavioural model for quantifying flood warning effectiveness. J .Flood Risk Manage. 4, 23–32.
Moore, R.J., Clarke, R.T., 1981. A distribution function approach to rainfall runoff modeling. Water Resour. Res. 17, 1367–1382.
Muluye, G.Y., 2011. Implications of medium-range numerical weather model output in hydrologic applications: Assessment of skill and economic value. J. Hydrol. 400, 448–464.
Murphy, A.H., Ehrendorfer, M., 1987. On the relationship between the accuracy and value of forecasts in the cost–loss ratio situation. Weather Forecasting 2, 243–251.
Nash, J.E., Sutcliffe, I., 1970. River flow forecasting through conceptual models. Part 1 – a discussion of principles. J. Hydrol. 10, 282–290.
Nielsen, S.A., Hansen, E., 1973. Numerical simulation of the rainfall-runoff process on a daily basis. Nord. Hydrol. 4, 171–190.
O'Connell, P.E., Nash, J.E., Farrell, J.P., 1970. River flow forecasting through conceptual models, Part ii – the brosna catchment at ferbane. J. Hydrol. 10, 317–329.
Oudin, L., Michel, C., Anctil, F., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 – can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? J. Hydrol. 303, 275–289.
Pagano, T.C., Wood, A.W., Ramos, M.H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., Verkade, J.S., 2014. Challenges of operational river forecasting. J. Hydrometeorol. 15, 1692–1707.
Pappenberger, F., Cloke, H.L., Parker, D.J., Wetterhall, F., Richardson, D.S., Thielen, J., 2015. The monetary benefit of early flood warnings in Europe. Environ. Sci. Policy 51, 278–291.
Parker, D., 1991. The damage-reducing effects of flood warnings. Report, Flood Hazard Research Centre, Middlesex University.
Parker, D.J., Tunstall, S.M., McCarthy, S., 2007. New insights into the benefits of flood warnings: Results from a household survey in england and wales. Environ. Hazards 7, 193–210.
Perrels, A., Frei, T., Espejo, F., Jamin, L., Thomalla, A., 2013. Socio-economic benefits of weather and climate services in Europe. Adv. Sci. Res. 10, 65–70.
Perrin, C., 2000. Vers une amélioration d'un modèle global pluie-débit. Thesis.
Perrin, C., Michel, C., Andreassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289.
Priest, S.J., Parker, D.J., Tapsell, S.M., 2011. Modelling the potential damage-reducing benefits of flood warnings using european cases. Environ. Hazards-Human Policy Dimensions 10, 101–120.
Ramos, M.-H., Mathevet, T., Thielen, J., Pappenberger, F., 2010. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Meteorolog. Appl. 17, 223–235.
Reggiani, P., Renner, M., Weerts, A.H., van Gelder, P., 2009. Uncertainty assessment via bayesian revision of ensemble streamflow prediction in the operational river rhine forecasting system. Water Resour. Res. 45, 1–14.
Richardson, D.S., 2000. Skill and relative economic value of the ecmwf ensemble prediction system. Q. J. R. Meteorolog. Soc. 126, 649–667.
Richardson, D.S., 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. Q. J. R. Meteorolog. Soc. 127, 2473–2489.
Rogers, D., Tsirkunov, V., 2010. Costs and benefits of early warning systems in global assessment report on disaster risk reduction Report. United Nations International Strategy for Disaster Reduction (UNISDR) and World Bank.
Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrol. Earth Syst. Sci. 11, 725–737.
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.
Salamon, P., Feyen, L., 2010. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. Water Resour. Res. 46, 1–20.
Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrol. Earth Syst. Sci. 16, 1171–1189.
Stephens, E., Cloke, H., 2014. Improving flood forecasts for better flood preparedness in the uk (and beyond). Geogr. J. 180, 310–316.
Sugawara, M., 1979. Automatic calibration of the tank model. Hydrol. Sci. 24, 375–388.
Thiboult, A., Anctil, F., 2015. Assessment of a multimodel ensemble against an operational hydrological forecasting system. Can Water Resour J 40, 272–284.
Thiboult, A., Anctil, F., 2015. On the difficulty to optimally implement the ensemble kalman filter: an experiment based on many hydrological models and catchments. J. Hydrol. 529, 1147–1160.
Thiboult, A., Anctil, F., Boucher, M.-A., 2016. Accounting for three sources of uncertainty in ensemble hydrological forecasting. Hydrol. Earth Syst. Sci. 20, 1809–1825.
Thielen, J., Bartholmes, J., Ramos, M.H., de Roo, A., 2009. The european flood alert system - part 1: concept and development. Hydrol. Earth Syst. Sci. 13, 125–140.
Thiery, D., 1982. Utilisation d'un modèle global pour identifier sur un niveau piézométrique des influences multiples dues à diverses activités humaines. In: Improvement of methods of long term prediction of variations in groundwater resources and regimes due to human activity. vol. 136. pp. 71–77.
Thornthwaite, C.W., Mather, J.R., 1955. The water balance Report. Drexel Institute of Climatology.
Valery, A., Andreassian, V., Perrin, C., 2014. 'as simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? part 1 - comparison of six snow accounting routines on 380 catchments. J. Hydrol. 517, 1166–1175.
Van den Bergh, J., Roulin, E., 2010. Hydrological ensemble prediction and verification for the meuse and scheldt basins. Atmos. Sci. Lett. 11, 64–71.
Verkade, J.S., Werner, M.G.F., 2011. Estimating the benefits of single value and probability forecasting for flood warning. Hydrol. Earth Syst. Sci. 15, 3751–3765.
Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5, 13–26.
Warmerdam, P.M., Kole, J., Chormanski, J., 1997. Modelling rainfall-runoff processes in the hupselse beek research basin. In: IHP-V, Technical Documents in Hydrology. pp. 155–160.
Weerts, A.H., Winsemius, H.C., Verkade, J.S., 2011. Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (England and wales). Hydrol. Earth Syst. Sci. 15, 255–265.
Zhao, R.J., Zuang, Y.L., Fang, L.R., Liu, X.R., Zhang, Q., 1980. The xinanjiang model. IAHS Publications 129, 351–356.
Zhu, Y.J., Toth, Z., Wobus, R., Richardson, D., Mylne, K., 2002. The economic value of ensemble-based weather forecasts. Bull. Am. Meteorol. Soc. 83, 73–83.