University of Reading

Estimation of uncertainty in flood forecasts - a comparison of methods

Article

Accepted Version

Estimation of uncertainty in flood forecasts - a comparison of methods

Leonore Boelee, Darren M Lumbroso

HR Wallingford, Howbery Park, Wallingford, Oxfordshire, UK, OX10 8BA

Paul Samuels

Cincinnati Children's Hospital
Anesthesiology
Cincinnati
Ohio
United States

Hannah Cloke

Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB United Kingdom

Abstract

The scientific literature has many methods for estimating uncertainty, however, there is a lack of information about the characteristics, merits and limitations of the individual methods, particularly for making decisions in practice. This paper provides an overview of the different uncertainty methods for flood forecasting that are reported in literature, concentrating on two established approaches defined as the *ensemble* and the *statistical* approach. Owing to the variety of flood forecasting and warning systems in operation, the question 'which uncertainty method is most suitable for which application' is difficult to answer readily. The paper aims to assist practitioners in understanding how to match an uncertainty quantification method to their particular application using two flood forecasting system case studies in Belgium and Canada. These two specific applications of uncertainty estimation from the literature are compared, illustrating statistical and ensemble methods, and indicating the information and output that these two types of methods offer. The advantages, disadvantages and application of the two different types of method are identified. Although there is no one 'best' uncertainty method to fit all forecasting systems, this review helps to explain the current commonly used methods from the available literature for the non-specialist.

Keywords: Flood forecasting, uncertainty, ensembles, probabilistic.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jfr3.12516

1. Introduction

Quantification of uncertainty in flood forecasts is moving from the realms of scientific research into practice. This potentially provides more reliable forecast information, a greater wealth of forecast information and longer lead times. There has been a move towards the use of ensemble Numerical Weather Prediction (NWP) models to provide meteorological forecast ensembles for flood forecasting systems. The use of NWPs, allows the uncertainty of the meteorological forecasts that drive the flood forecasts to be assessed and forecast lead times to be extended (Cloke and Pappenberger, 2009; Smith *et al.*, 2016; HEPEX, 2017). Although the meteorological forecast is an important source of uncertainty, uncertainty is present in all components of a flood forecasting system (Pappenberger et al., 2005). The typical component of a flood forecasting system are represented by the blue boxes shown in Figure 1. A growing range of techniques is available in the flood forecasting literature for quantifying uncertainty, sensitivity, risk and decision analysis. However, no well-accepted guidelines exist on implementing these principles and techniques for the multiple sources of uncertainty affecting flood forecasting systems (Zappa et al., 2010; Liu and Gupta, 2007). A lack of coherent terminology or systematic approaches means that it is difficult (or perhaps impossible) to assess the characteristics and limitations of individual methods and select the most appropriate method for any particular case (Montanari, 2007) particularly for those working outside -academia. The process of selecting the most suitable method to predict uncertainty is no different in principle from selecting the most suitable model for predicting a flood. This is described as follows: ‘The essential question is not which model (method) is the more suitable for flood forecasting, but rather what type of information a decision maker needs and how they can proficiently use it to produce good and, possibly, more reliable decisions’ (Todini 2017).

For flood forecasting practitioners the available literature on uncertainty methods for flood forecasting is often opaque which results in ineffective use of the different uncertainty methods in practice. In addition, it is often not clear what questions concerning uncertainty these different methods can and cannot answer. This means that it can be challenging to identify where the strengths and weakness lie when applying the methods operationally. However, the question, ‘Which uncertainty method is most suitable for which application?’ is too broad to answer, owing to the many different types of flood forecasting systems. Hence, this paper takes an alternative approach to help clarify the quantification of uncertainty for flood forecasting practitioners by providing an overview of what information an uncertainty method can provide into the flood warning process.

This paper concentrates on describing what information two established uncertainty methods, based on the use of statistics and ensembles, can provide to flood forecasting processes, by discussing two operational fluvial flood forecasting systems in Belgium and Canada. The objective of this paper is to identify the main assumptions behind the methods, their strengths and drawbacks in a way that is easy to follow for specialist and non-specialist alike. This paper is aimed at an audience that does not necessarily have prior knowledge on the topic of uncertainty and therefore it includes definitions on terminology used frequently in this and other papers on flood forecasting uncertainty in Table 1 (italics indicate the term is in Table 1). The paper starts with a short overview of the terminology used for uncertainty and the approaches adopted by flood forecasting systems. It concludes with identifying where the remaining challenges and opportunities lie in the application of uncertainty methods to flood forecasting systems.

2. Overview of uncertainty

2.1. Definitions of uncertainty

Uncertainty results from the lack of knowledge or the inability to accurately measure or calculate an observed value, which can lead to differences between the modelled and its ‘true’ value of a variable (Gouldby and Samuels, 2009). Two types of uncertainty can be defined: aleatory and epistemic. Aleatory uncertainty is the uncertainty due to the natural variability of the physical world and reflects the inherent randomness in nature, whereas, epistemic uncertainty is the uncertainty due to a lack of knowledge of the physical world and a lack of ability to measure and model it (Li et al., 2013). This division can be useful as it distinguishes which uncertainties can be reduced and which cannot (Der Kiureghian and Ditlevsen, 2007).

The literature on uncertainty in flood forecasting frequently uses the terminology ‘predictive uncertainty’ or ‘predicting the uncertainty’, examples include the work of (Palmer, 2000; Todini, 2008; Weerts et al., 2011; Zappa et al., 2011; Van Steenbergen and Willems, 2015). Todini, (2008) defines predictive uncertainty as “the probability of any future (real) value, conditional upon all the knowledge and information, available up to the present”. The popularity of the term “predictive uncertainty” in flood forecasting research is because it emphasizes that it is the uncertainty around the prediction which is being described or quantified, rather than “validation uncertainty” or “model uncertainty”; which are defined as: “the ability of a model to reproduce reality” (Todini, 2008; Klein et al., 2016).

Table 1: Some definitions commonly used in uncertainty quantification and flood forecasting and/or used in this paper. Italics alerts the reader that a term is defined in this table.

Term	Definition
Correct alarm (HIT)	When both modelled and observed values exceed a warning threshold.
Correct alarm ratio (CAR)	The ratio between the correct alarms (HIT) and the summation of the HITs and False Alarms (FA). A forecasting the model exhibits a high forecast skill when CAR values are close to one.
Correct rejection (CR)	When both simulated and observed value are below the warning threshold.
Decision makers	Anybody aiming to interpret and use a flood forecast to disseminate a flood warning or to take other actions to mitigate and reduce flood risk from an imminent event.
Deterministic forecast	A forecast that provides definite information as a single forecast. In meteorology a deterministic forecasting model is often run at a higher spatial resolution than a probabilistic model.
Ensemble forecasting	Ensemble forecasting provides a set of many plausible forecasts rather than providing a single deterministic forecast of future conditions, ensemble forecasting is common in NWP.
Ensemble Kalman filter (EnKF)	Computational algorithm that uses an ensemble in processing measurements to find an optimum estimation of the past, present or future states of a system.
Ensemble spread	The total area of variation between the most upper and lowest ensemble. A wide spread indicates a high uncertainty and low predictability.
Ensemble members	Individual forecasts within an ensemble.
Event based models	Models simulating a limited period of time using observed rainfall from the past or a hypothetical event of some estimated probability of occurrence.
False alarm (FA)	When the forecast value exceeds a warning threshold but the observed value does not.

Term	Definition
Feasible model space	The set of all the plausible values required to run a model that can be found by using a finite combination of all the possible model initial conditions, parameters and/or boundaries conditions. This will often lead to a great number of values which can reduced by sampling using for example the Monto Carlo technique.
Forecast accuracy	The degree to which the forecast variable conforms to its observed value.
Forecast skill	How much better a forecast is compared to the long term average on that day (climatology).
Heteroscedasticity	This refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it.
Initial conditions	The system state at the start of the simulation. For example for a hydrological model these can be the initial state of the soil moisture, snow cover, water level or river flow. Also referred to as antecedent conditions or model states.
Lead time	The time between the forecast (or the issued warning) and the arrival of the predicted flood , or the length of time into the future that is forecasted.
Miss rate (MR)	The ratio between the missed alarms (MIS) and the summation of the MIS and correct rejections (CR). The model shows high forecast skill for a CAR value close to zero.
Missed alarm (MIS)	When the actual value exceeds the warning threshold, but the simulated one does not.
Model spin-up	When the initial conditions are unknown, the model can be run for a period of time after which the initial conditions are assumed to no longer significantly impact the results. This time period is referred to as the model spin-up or model warm up.
Model residual	The difference between the modelled value and the observations, sometimes termed ‘model error’.
Predictive uncertainty	The probability of any future (real) value, conditional upon all the knowledge and information, available up to the present, see Todini, (2008).
Probabilistic forecast	When a forecast includes the associated probability of the event occurring.
Reference Climatology	The 30 year average hydrological or meteorological conditions for the calendar period in question.
Under-dispersed ensemble	When the width of the probability density of the forecast ensemble is narrower than the observations. This means that the ensemble members are too similar to each other and different from the observations.

2.2. Sources of uncertainty

Uncertainty affects each element in a flood forecasting system. Which sources of uncertainty are pertinent will depend on what elements the forecasting system contains, as shown in Figure 1. For example, if a hydraulic model is part of the chain which provides flood level and inundation extent, additional uncertainty sources are present compared to a system which uses a rainfall-runoff model to produce river flows alone or a simple level to level correlation. This means that in the literature the named sources affecting a forecast system can vary. In the literature the most commonly discussed sources of uncertainty for flood forecasting systems include:

- Uncertainty inherent in the meteorological forecast: Precipitation is often considered one of the most important atmospheric inputs into a flood forecasting system, especially in catchments without snowmelt processes. NWP is commonly used to forecast precipitation. The uncertainty in NWP forecasts are related to the initial conditions, boundary conditions and model uncertainty, which are assessed using an ensemble. Ensembles comprise multiple weather and climate prediction models with explicit perturbation of initial conditions and model formulations (Palmer, 2000). Challenges in using precipitation forecasts for flood forecasting systems include: the scale of the atmospheric model does not necessarily match the hydrological model (uncertainty due to downscaling is not considered further in this paper, for more information see (Maraun et al., 2010; Schoof, 2013; Gutmann et al., 2014)); increasing uncertainty in precipitation forecasts for rare events; increasing uncertainty in precipitation forecasts with increasing forecast lead time and uncertainty related to convective precipitation which cannot be resolved at 10-40km grid scales, a particular issue for medium range predictions (5 to 10 days) (Pappenberger *et al.*, 2005; Rossa *et al.*, 2011). For a comprehensive review on uncertainty of precipitation forecasting see (Palmer, 2000; Rossa *et al.*, 2011; Liguori and Rico-Ramirez, 2014).
- Uncertainty from measurement and observations: The calibration and verification of flood forecasting system uses observations, which themselves are uncertain (Gotzinger and Bardossy, 2007). Observations are subject to random and systematic errors that can vary over time. The spatial and temporal characteristics of the observations do not always directly correspond to the modelled fluxes and storages leading to uncertainties added during the processing of observation (Juston et al., 2013). For example, interpolation can lead to errors and issues with capturing the spatial variability (Gotzinger and Bardossy, 2007) and using rating curves, which relate water levels to flows, are a major source of uncertainty in discharge estimations (Di Baldassarre and Montanari, 2009; McMillan et al., 2012). For a complete overview of uncertainty due to measurements and remote sensing data see (McMillan et al., 2012; Li et al., 2016; Li et al., 2016).
- Uncertainty due to initial conditions: Uncertainty due to initial conditions relates to the uncertainty of the land surface state, including the soil moisture, snow cover, initial state of the river and other waterbodies in the catchment (Madsen and Skotner, 2005; Gotzinger and Bardossy, 2007; Li et al., 2009). Land surface state measurements are often not in proportion to the heterogeneity of the land surface and this is a source of uncertainty. For example soil moisture is often a single point measurement which will be spread over to the modelled catchments or grid (Beven and Binley, 2014).
- Uncertainty due to hydrological and hydraulic models being unable to fully represent processes: The inherent simplifications of the model in order to represent the more complex real system leads to uncertainty. Model structure uncertainty refers to the uncertainty of the represented processes, the chosen representations (e.g. St. Venant or kinematic wave equation for channel routing) and the spatiotemporal scales used in the model (Smith et al., 2016). An example of this is the use of polygons or grids to represent catchments, this will lead to uncertainty due to the physical processes often occurring on smaller scales than the model elements (Gotzinger and Bardossy, 2007).
- Uncertainty due to model parameters: The estimation or calibration processes of parameters in models inevitably leads to uncertainty. Catchment characteristics have natural variability, which leads to local spatial heterogeneities and non-stationarities in the catchments affect the parameters, making them difficult to estimate effectively (Gupta et al., 2003). A second issue affecting the uncertainty of model parameters is that the performance of a calibrated model in prediction is not necessarily as good as during calibration (Beven and Binley, 2014). This can be caused by future conditions inducing a different type or range of responses beyond the models calibrated range (Beven, 2012). For a comprehensive overview on parameter uncertainty the reader is referred to (Beven and Freer, 2001; Vrugt et al., 2003).

The above sources of uncertainty are essentially epistemic in character, although arguably they all have aleatory components. These sources of uncertainty are often combined in: input (observations, downscaling and NWP), *initial conditions* and modelling uncertainty (model structure and/or model parameter uncertainty)

(Liu and Gupta, 2007; Zappa et al., 2011; Van Steenbergen and Willems, 2015; Klein et al., 2016; Thiboult et al., 2016).

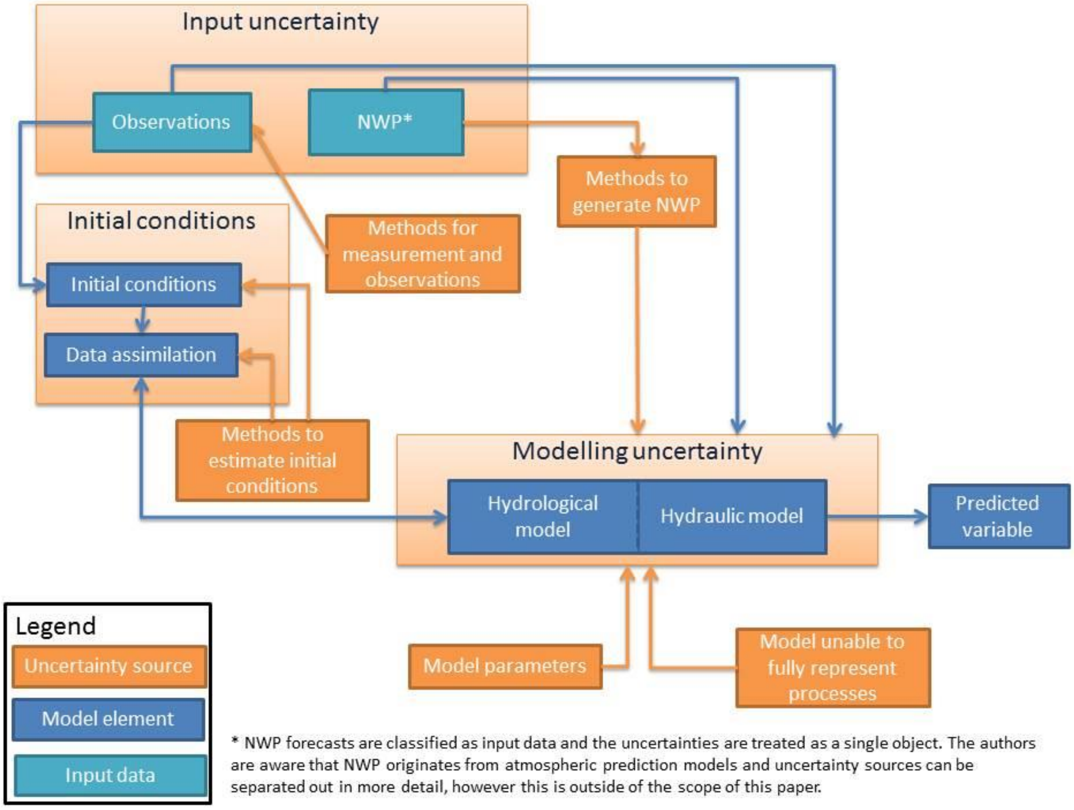

Figure 1: Model elements and sources of uncertainty in a typical flood forecasting system

2.3. Uncertainty methods in flood forecasting

In quantifying flood forecasting uncertainty two different philosophies can be identified. The first is using the statistical analysis between forecast and observed values as a measure of uncertainty. Methods in this philosophy are referred to as ‘statistical methods’. The second is philosophy is using a set of plausible forecasts as a measure of uncertainty. Methods within this philosophy are referred to as ‘ensembles methods’. Some uncertainty methods use a combination of both of these two philosophies.

The statistical methods can be referred to as post processors. They calculate the *model residual* and are based on the assumption that the model uncertainty from the past is representative of the uncertainty in the future. The information the statistical method provides is an estimation of the uncertainty of a specific water level/discharge for a specific lead time. The question this estimation procedure answers is: ‘What is the probability of the forecasts being accurate, based on past performance?’ Statistical methods range in complexity and in their assumptions. Some methods like the Hydrological Uncertainty Processor (HUP) (Krzysztofowicz, 1999) make direct assumptions on the distribution of the model residual, where other methods like Quantile Regression (QR) (Koenker, 2005) avoid this, but make other assumptions in order to calculate the quantiles (Wani et al., 2017). Finding a reliable statistical methods to use model residuals to represent uncertainty remains a mathematical and theoretical challenge.

Ensembles are generally created by combining different runs, where each run is within the *feasible model space* of the model structure, model parameters and forcing data. The ensembles method assumes that it is possible to define the model structure and parameter space which is representative for the predictive uncertainty. The information the ensemble method provides is a measure of the spread of the forecast based on the lack of knowledge on the models processes, parameters and/or initial conditions (Todini, 2017). The question the method can answer is: What is the likely spread of the forecast given the known lack of knowledge on the model structure, parameter and/or initial conditions of the catchment, river and/or atmosphere? Finding a reliable methods to create and sample from model space is a mathematical and theoretical challenge. Methods to create ensembles reflecting uncertainty in flood forecasting models focus on one or more uncertainty sources (Boucher et al., 2012). For example the Shuffled complex evaluation metropolis algorithm (Vrugt et al., 2003) creates an ensemble of model parameters, the *Ensemble Kalman Filter* (Evensen, 2003) creates an ensemble of *initial conditions* and the concept of using a multi-model ensemble treats model structural uncertainty (Thiboult et al., 2016; Todini, 2017).

There are uncertainty methods that use a combination of both philosophies. For example (Hemri et al., 2013) use statistical methods to improve the *hydrometeorological ensemble* which has been produced by rerunning a hydrological model with a meteorological ensemble.

An overview of methods capturing uncertainty that are applied to flood forecasting systems is presented in a table in the supplementary material and as summary in **Error! Reference source not found.**. The objective of the table is not to be complete, but to give a representative indication of the variety of methods available to flood forecasters. It is shown that uncertainty methods are able to deal with parameter uncertainty, uncertainty due to the meteorological forecast, model structure uncertainty, uncertainty in gauged flow data and 'total' uncertainty as an aggregate. Most methods have been applied to hydrological models, although there are examples of hydraulic models and inundation models. The catchments that the uncertainty methods have been applied to vary both in size and location.

3. Application of a statistical method

This section provides an analysis of the character of a statistical method applied to forecast *model residuals*, using as an example the Non Parametric Databased Approach (NPDA) in the Belgian case reported by (Van Steenbergen et al., 2012).

3.1. Application description

Van Steenbergen et al., (2012) describe the forecasting system at the Flanders Hydraulics Research Centre (FHRC) (FHR, 2017) for navigable rivers which provides *deterministic forecasts* several times a day with a 48 lead time for the main river in Flanders including the Yser, Dender and Demer, more information on the forecasting systems for these rivers is in Figure 4 and Table 2. During extreme flood events the system performed less well than hoped. For some catchments forecasts did not meet the acceptable level at a maximum relative error of 10%. The uncertainty in the forecasts is not surprising or indeed unique to this system. Any flood forecasting system that simulates complex hydrological and/or hydraulic processes will do so with a degree of uncertainty (Leedal et al., 2010; Pappenberger et al., 2005). To allow end users, flood managers and other *decision makers* to account for this uncertainty, transitioning to *probabilistic forecasts* was explored by applying Non Parametric Databased Approach (NPDA). The NPDA uses a statistical analysis of the *model residual*. The residuals are divided into classes where the distribution of the forecast residuals is assumed the same. The percentiles are calculated per class for the different lead times and populated into a three dimensional matrix. This matrix is used as a lookup table to provide a forecast with confidence intervals. The schematisation of the NPDA method is shown in Figure 5. FHRC explored the NPDA due to its stability and speed, an ensemble approach was rejected due to its excessive computational needs.

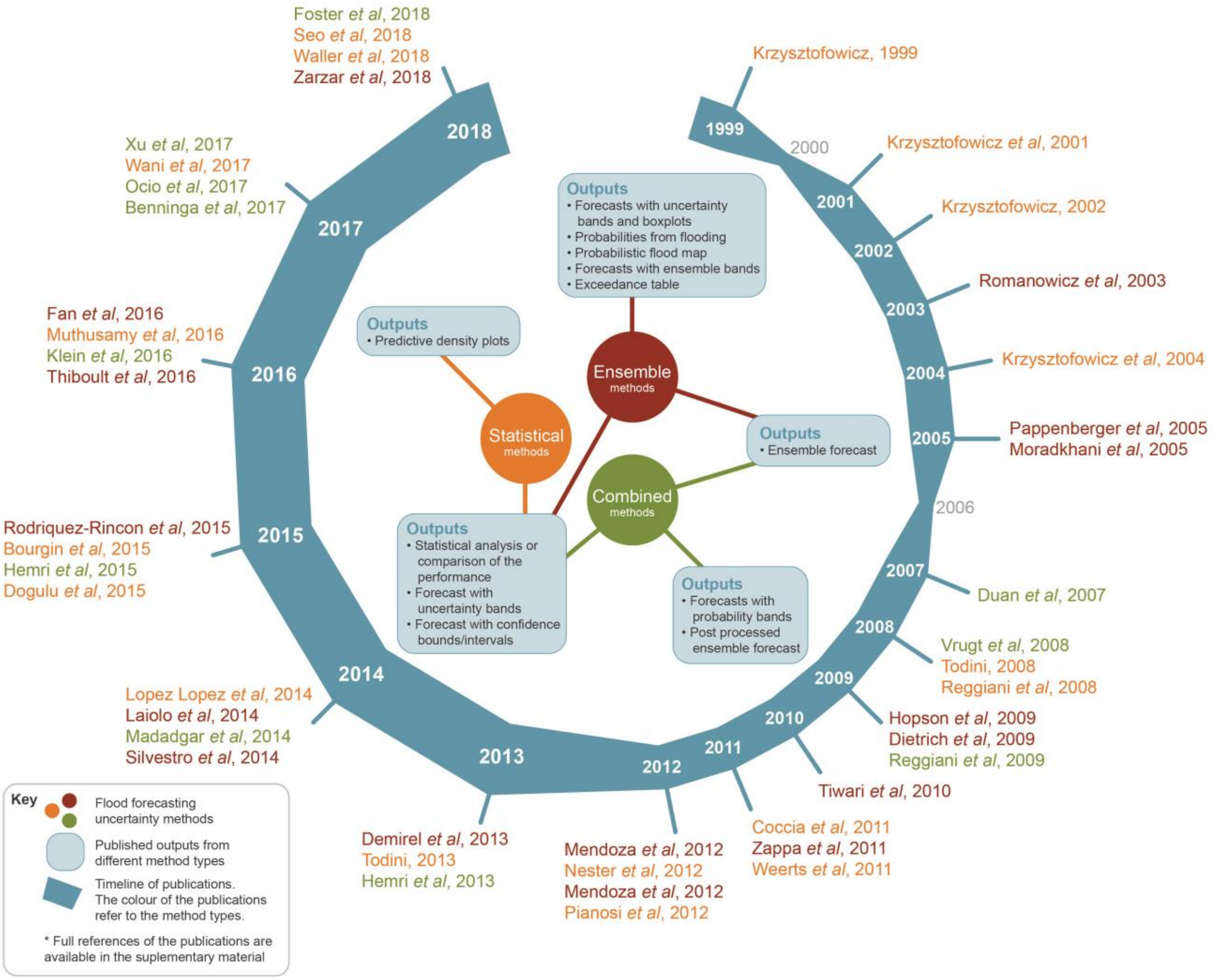

Figure 2: Chronological overview of available methods for uncertainty estimation in flood forecasting based the table available in the supplementary material.

Table 2: Characteristics of the application of the Non Parametric Databased Approach, a statistical method.

Location	Flanders, Belgium. The rivers Yser, Dender and Demer, overview in Figure 4.
Forecast centre	Flanders Hydraulics Research Centre for Navigable Rivers.
Catchment mean annual rainfall	700 to 800mm per year.
Average annual flows	2 m^3/s to 15 m^3/s.
Catchment area	1,101km^2 (Heylen, 1997) to 2,275 km^2 (Cauwenberghs and Maeghe, 2007)
Catchment description	Catchments with diverse land use including arable, urban and forest. Rivers including flow regulation structures in the form of hydraulic gates and sluices.
Forecast length	48 hours

Models used in the flood forecasting system	Lumped conceptual rainfall-runoff model for upstream catchments, hydrodynamic models for the main rivers and data-assimilation for real-time updating.
Flood plain representation	Quasi-two dimensional i.e. (the floodplain was schematized into one dimensional river branches linked to the main river by spills).
Data for mode setup	Field survey data at 50 m intervals, 5x5 m with ±0.1 m vertical resolution digital elevation model (for flood plain representation).
Data for calibration	Observed rainfall, evaporation, level and flow time series. Catchment average rainfall generated using Thiessen polygon approach. For the ungauged parameters of the neighbouring catchment are used.

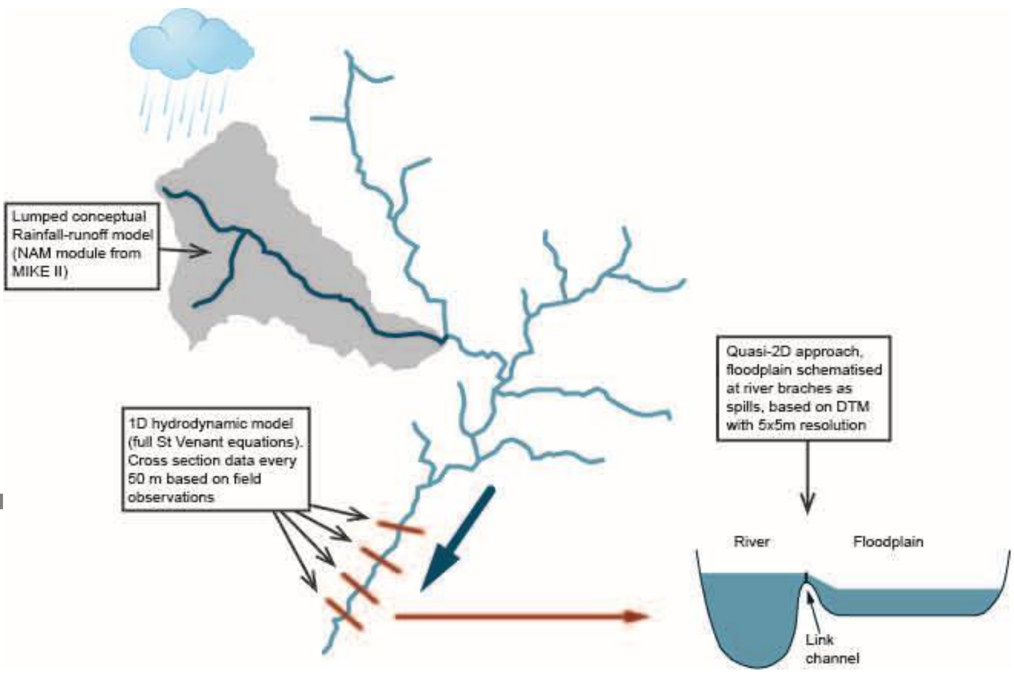

Figure 3: Schematisation of the Flood Forecasting models at FHRC (figure produced by L. Boelee)

3.2. Performance of a statistical approach

To assess the performance of the *probabilistic forecast* (Van Steenbergen et al., 2012) compared the exceedance of alarm levels of the deterministic with the probabilistic forecasts for three catchments. The probabilistic forecast was optimised using the *Correct Alarm Ratio* (CAR) and *Miss Rate* (MR). The CAR-MR is a value between 0 and 1, with 1 being a perfect forecast. All four catchments show in increase in the CAR-MR score. Two out of the four catchments had a high CAR-MR for the *deterministic forecasts*. Using the probabilistic forecast for these catchments showed minor increase in the CAR-MR score (0.01 and 0.02) when compared to the deterministic forecasts. For the other two catchments where the performance of the deterministic forecast was lower according to the CAR-MR scores, the use of probabilistic forecast showed an improved performance with the CAR-MR score increasing with 0.1 for one catchment and a minor increase in the CAR-MR score 0.02 for the other. Producing probabilistic forecasts using the non-parametric data approach can be beneficial for the *forecast accuracy*.

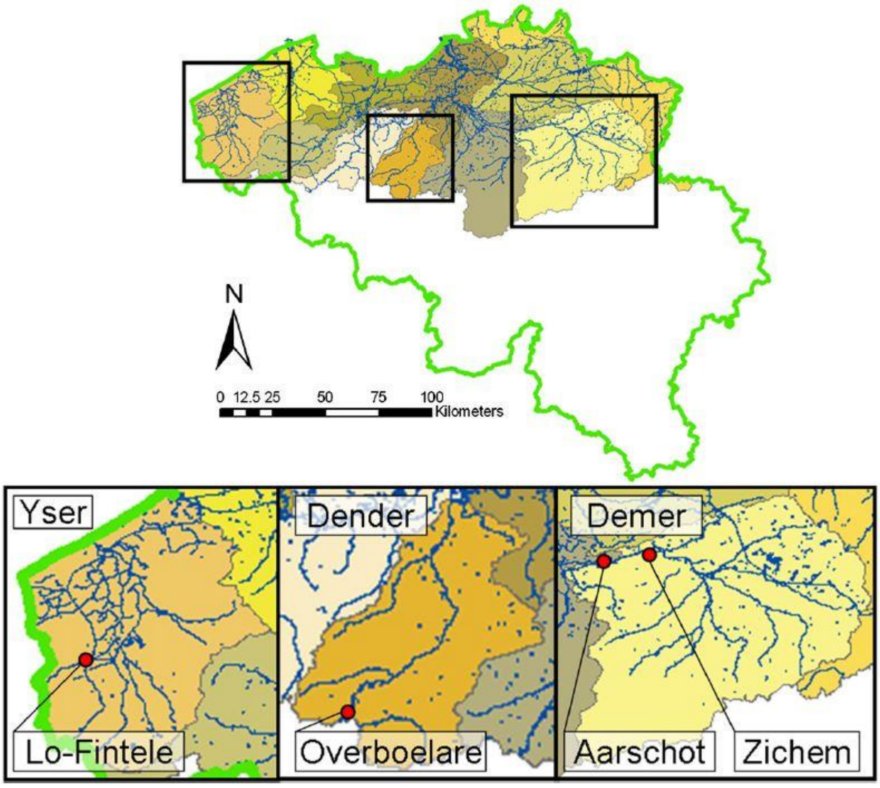

Figure 4: Catchment locations of application 1, the NDPA

Source: Reproduced from (Van Steenbergen et al., 2012) with permission of the authors.

3.3. Advantages and disadvantages

The advantages of using a *probabilistic forecast* generated by the NPDA is that there is an improved forecast performance. This is expressed quantitatively by the increase shown in the CAR-MR scores for the probabilistic forecasts when compared with the deterministic ones. The method is straightforward to apply to a new or existing system and requires modest computational resources. The speed and simplicity of the method means it can be combined with flood forecasting models that are run at a high frequency and where the hydrological and/or hydraulic models have longer run times. The method can be used to estimate 'total' uncertainty of a forecast, aggregating initial condition uncertainty, past meteorological uncertainty and uncertainty from the hydrological and hydraulic models. The method's strength lies in applying it to a forecast system where probabilistic forecasts are required, but there is limited budget and computational resources, or a forecast system where it is undesirable to change the current deterministic setup.

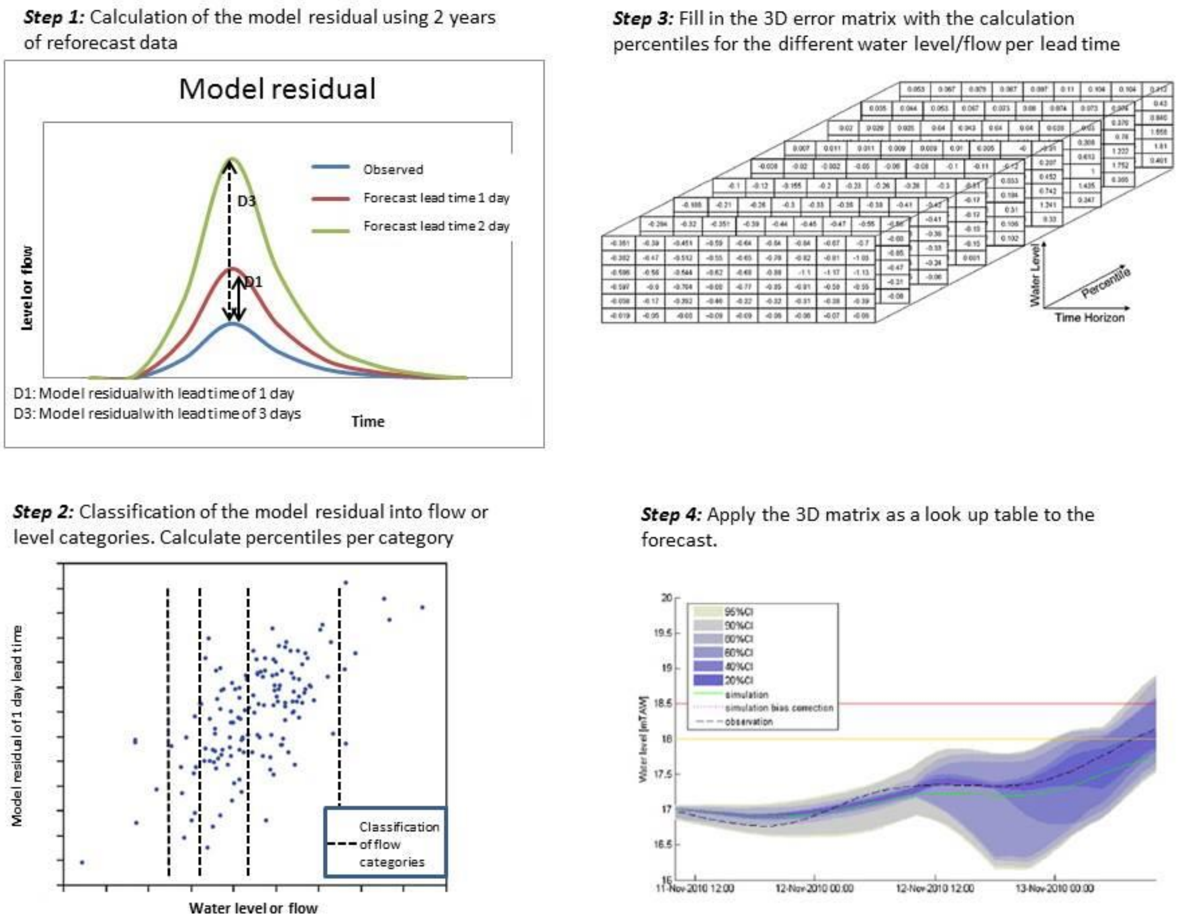

Figure 5: Schematisation of the NPDA

Source: Figures at step 3 and 4 reproduced from (Van Steenbergen et al., 2012) with permission from the authors. Figures at step 1 and 2 produced by L.Boelee.

The disadvantages of this approach is that it relies heavily on observed streamflow data and requires regular updates with new data, six monthly in the Belgian case. Owing to the databased aspect of the method the uncertainty matrix is directly linked to the gauged location and can only be reliably used there. The transferability of the uncertainty matrix to an ungauged location was not explored. Van Steenbergen (2012) uses two years of forecasts and observed data for the setup of the method, this could be a disadvantage for new systems. Another drawback of the method is that the uncertainty is not captured equally across the water level/discharge spectrum. Classes in which the percentiles are calculated will require a minimum number of data points. For more extreme flows there will be fewer model residuals and therefore the water level or flow class will need to be broadened, leading to less reliable percentile calculations. Extreme values that have not occurred before will assume the percentiles in the most extreme classes are representative.

4. Application of an ensemble approach

This section provides an analysis of the character of a typical ensemble forecasting approach, using as an example the three-hourly forecasts in the Canadian case reported by (Thiboult et al., 2016). This operational system uses the ensemble meteorological forecasts available from the European Centre for Medium-Range Weather Forecasts (ECMWF).

4.1. Application description

Thiboult et al., (2016) describe the flood forecasting system of Québec (Table 3) which issues five days of three hourly stream flow forecasts to municipal water managers and five daily forecasts to the public. Currently a statistical method assesses uncertainty in 10 river basins, more details in (Centre d'expertise hydrique du Québec and MDE, 2017) (Matte et al., 2017). With the aim of seeking more accuracy and reliability of the streamflow forecasts and improvement in the current estimation of uncertainty, an experiment was performed to disaggregate the sources of uncertainty. Three different ensembles were used to capture three types of uncertainty; meteorological uncertainty, *initial conditions* uncertainty and structural uncertainty of the hydrological models, shown in Figure 7. The assumption when using an ensemble is that the space from which the ensembles are sampled has been defined such that the *ensemble spread* is representative for the predictive uncertainty. For the meteorological uncertainty the 50 member ensemble from ECMWF was downscaled and for the initial conditions an *ensemble Kalman Filter* (EnKF) was used, producing a 50 member ensemble (Thiboult and Anctil, 2015a). The EnKF has been setup to assimilate the gauged stream flow data and has been optimised with all the hydrological models, more details on the EnKF see (Thiboult and Anctil, 2015b). To represent structural uncertainty of the hydrological model an ensemble of 20 members was used, by selecting 20 hydrological models of varying complexity (Seiller et al., 2012). The system is setup flexibly allowing the different ensembles to be turned 'off' or 'on', represented by the dotted arrows in Figure 7. In this experiment the evapotranspiration and snowmelt schemes were not varied, meaning that the model structural uncertainty in these component is not included in the predictive uncertainty.

Table 3: Characteristics of the application the ensemble method.

Location	Québec, Canada. 20 catchments, overview in Figure 6.
Forecast centre	Ministry of sustainable development, the environment and climate change (Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques).
Catchment mean annual rainfall	877mm to 1,412mm.
Average annual flows	8 to 300 m^3/s.
Catchment area	512km^2 to 15,342km^2.
Catchment description	Natural catchments without any influence of dams and structures.
Forecast length	Five day forecast operationally, for the uncertainty experiment a 9 day forecast.
Models used in the flood forecasting system	Operational forecasting system uses a semi-distributed physics-based hydrological model HYDROTEL (Fortin et al., 1995). The uncertainty experiment uses 20 lumped rainfall-runoff hydrological models (Seiller et al., 2012).

Data for calibration	10 years of data, four years of data were used for the *model spin-up* and two years of forecast were created using 2 years of meteorological forecasts. The data included temperature, precipitation, telemetered flow data, forecast temperature and forecast precipitation.

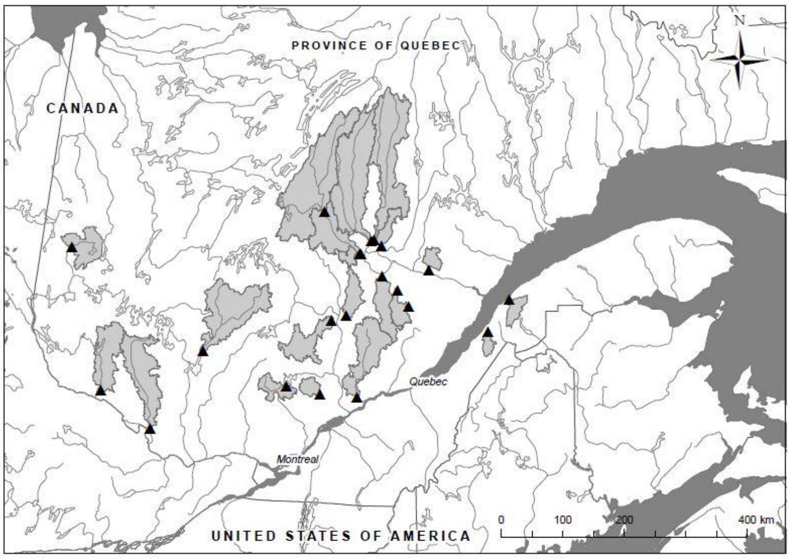

Figure 6: Catchment locations of application 2, the ensemble method

Source: Reproduced from (Thiboult et al., 2016) with permission from the authors

4.2. Performance of the ensembles approach

Thiboult et al., (2016) use different scores to assess the performance of the ensembles, which include comparing the *probabilistic forecast*, *deterministic forecast* and observations. The scores include the continuous ranked probability score, CRPS (Matheson and Winkler, 1976), reliability diagram (Stanski et al., 1989), mean absolute uncertainty of the reliability diagram which is the average distance between the forecast and observed frequencies over all quantiles of interest (Brochero et al., 2013; Thiboult et al., 2016) and Spread Skill Plot which takes the root mean squared error (RMSE) compared to the square root of average ensemble variance for which the spread should match the RMSE (Fortin et al., 2014). Cloke and Pappenberger (2008) provide more information about skill scores. The forecast improvement and representativeness of the uncertainty are assessed by these scores.

The results show the contributions of the three sources of uncertainty change when looking at different lead times and catchments. The meteorological uncertainty increases for longer lead times as is typical. With this specific probabilistic forecast (downscaled ECMWF) it improved forecast performance substantially for a nine day lead time, but only marginal improvements are visible for shorter lead times. It should be noted that for this particular case the ECMWF downscaled precipitation forecast is shown to have an *under-dispersed ensemble spread*. There are methods available that aim to improve the ensemble spread of the precipitation forecast (Verkade et al., 2013), however this was not part of Thiboult et al., (2016) research. Including hydrological structure uncertainty improved the forecasts substantially through the shorter lead time (up to three days). For day six there is a minor improvement and for day nine using the 20 member hydrological ensemble showed no more skill than the *reference climatology*. The *initial conditions* ensemble shows a significant improvement of skill for the short lead time (up to three days) and a minor improvement compared to the deterministic forecast for day six. For day nine using the 50 ensembles generated by the EnKF does not show more skill than the *reference climatology*. The hydrological ensembles and the ensemble of initial conditions overlap in the treatments of uncertainty sources and indeed show a similar forecast improvement.

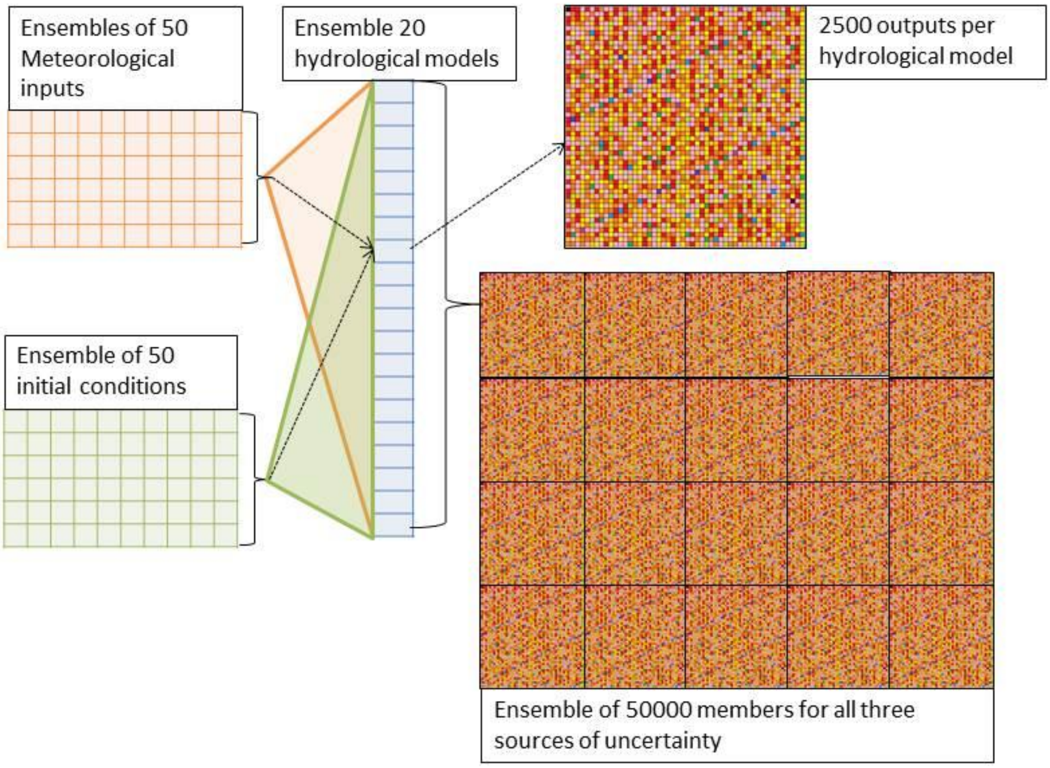

Figure 7: Overview of the ensemble approach in the Quebec catchments (figure produced by L. Boelee).

4.3. Advantages and disadvantages

The advantages of the ensemble method are that combining the EnKF, the meteorological ensembles and the hydrological multi model to generate a *probabilistic forecast* offers an improved *skill score* compared to the *deterministic forecast*, for all lead times. The approach allows the different sources of uncertainty to be assessed separately in an operational setting. This knowledge can be used to give *decision makers* information on the total uncertainty and on where that uncertainty is coming from, offering richer information for decisions. The approach does not relay on a specific gauged location and results with uncertainty estimation can be extracted anywhere in the modelled domain and can be applied to any forecasting system where probabilistic forecasts are required.

The disadvantages of the ensemble method are that creating an ensemble depends on the definition of the feasible model structure and parameter space. However, defining this space in itself is uncertain, which can lead to the ensemble being under or over dispersive and thus not representing the predictive uncertainty accurately. For some forecasting systems the total ensemble size of 50,000 members will be a significant drawback for use in practice, owing to the consequent required increase in computation resources, data management and fast running of models. Thiboult et al., (2016) found redundancy within the total ensemble, but effectively reducing the ensemble was not part of the research. Reducing the ensembles is not as straightforward as selecting the most dominant source of uncertainty due to the varying nature of the uncertainty across catchments and lead times,

5. Comparison and selection of the methods

Using either the ensemble or statistical method to generate probabilistic flood forecasts has shown an increase in the skill scores as described in Sections 3 and 4, compared to using the *deterministic forecast*. Both the ensemble method and the statistical method are based on a different set of assumptions and capture different aspects of uncertainty. The statistical method is based on the assumption that the past performance is representative of the uncertainty in the future and the ensemble method bases the uncertainty on the lack of knowledge on the models processes, parameters and/or initial conditions. In addition to these differences in the information that the methods capture, there are differences in the application, practicality and outputs of both methods. These differences are summarised in Table 4.

Table 4: Overview of the statistical and ensemble methods

	Statistical methods	**Ensembles**
Computational requirements, resources	Most methods have low computational requirements and resources.	Dependent on the size of the ensemble, however computational requirement and resources are likely to be higher than the statistical methods.
Application to an existing system	Can be added onto an existing flood forecasting chain as a post process.	Forecast system would need to be structured in a way that makes rerunning of models within of the forecasting system possible
Fixed location or whole domain	Can only be applied to locations with gauged stream flow or level data	Can be generated for the whole model domain
Sources of uncertainty	Captures the ‘total’ uncertainty	The uncertainty can be disaggregated per sources or targeted to a single source

Four practical steps represent a possible approach for selecting a method and applying uncertainty in a current or future flood forecasting system:

1. List the constraints of your current of future system; e.g. model type and run times, requirements of a pre-existing system, computational resources and data availability.
2. Identify how the uncertainty will be used and what the associated requirements that will result in. For example, does the uncertainty need to be broken down into components, is there a NWP ensemble that needs to be included etc.
3. Identify which type of method suits your application best: ensemble, statistical or combined, use Figure 8.
4. Use the table in the supplementary material to find literature on suitable methods for your method type and catchment type.

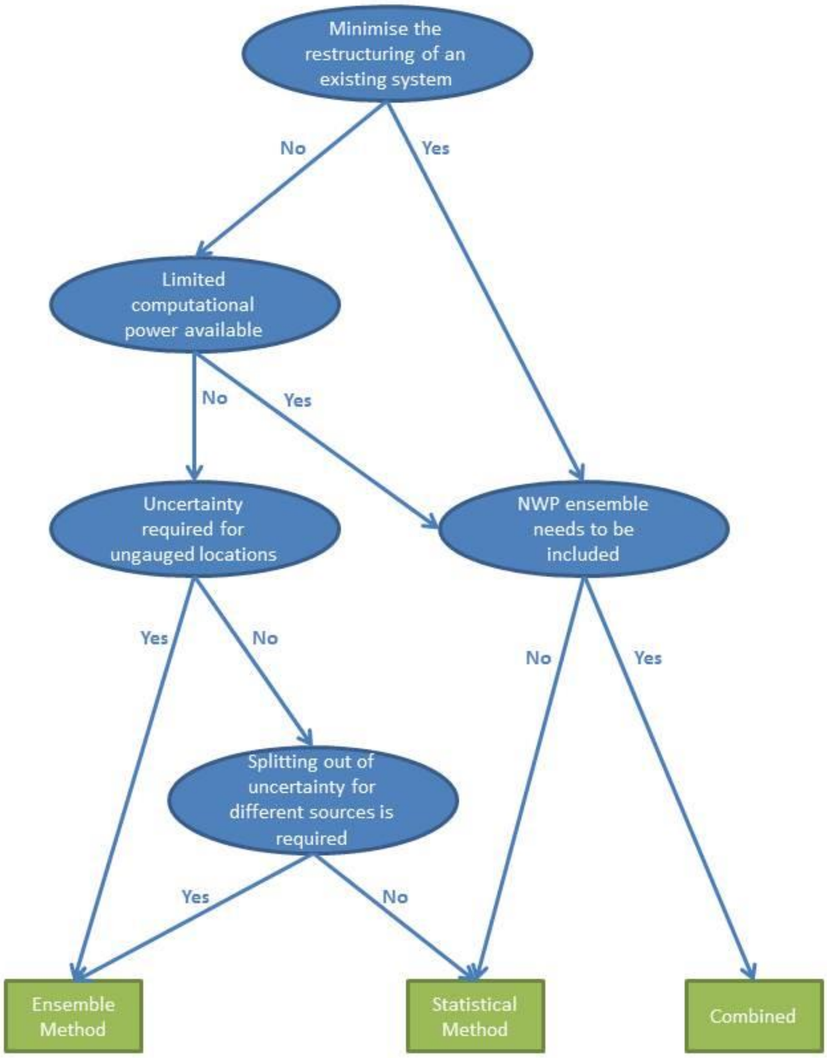

Figure 8: Selection of Methods flow chart

6. Discussion

Five challenges for using and researching uncertainty in flood forecasting have been identified during this review.

The **first challenge** is that there are many different definitions for uncertainty that are used within flood forecasting. There have been calls for a more **coherent terminology** (Montanari, 2007), however this has proved difficult to achieve. In the last decade there has been more consistency in using the term *predictive uncertainty* as defined by (Krzysztofowicz, 1999; Todini, 2008) to describe uncertainty in flood forecasting systems as opposed to uncertainty in *event based models*. However, there is little agreement on how predictive uncertainty is to be quantified, with (Todini, 2017) referring to uncertainty from an ensemble method as forecast sensitivity rather than uncertainty and (Matte et al., 2017) referring to a statistical methods as ‘dressing’ a deterministic model. Although finding agreement on terminology is not a trivial matter, greater clarity can be achieved by defining the technical terms within a piece of work as in the FLOODsite Language of Risk (Gouldby and Samuels, 2009) and in this paper in Table 1.

The **second challenge** is that both statistical and ensemble uncertainty quantification methods have **mathematical and theoretical challenges** remaining. Currently it is unclear what assumptions have been made in order to quantify uncertainty and what the consequences of these assumptions are, indicating a lack of knowledge about uncertainty estimation. Beven (2016) provides a comprehensive discussion about the theoretical challenges of uncertainty quantification from the perspective of their type (e.g. aleatory, epistemic). When using simulation and resampling techniques the challenges are how to create a reliable and practically sized ensemble. When using techniques based on statistical analysis of *model residuals* the challenges are how to deal with creating representative uncertainty bands for the future based on the historical model residual, this includes dealing with non-stationarity and *heteroscedasticity* of the residual. Research questions on the quantification of uncertainty that this review has brought forward are: How do the assumptions made when creating an ensemble affect the assessment of predictive uncertainty? And: What are the implications of assuming that the statistics of the historical model residuals are representative for the predictive uncertainty? This calls for more applied research testing different uncertainty quantification methods on different catchments and flood events in a comparable way.

The **third challenge** is that research on the **representativeness of the uncertainty spread** for prediction and extreme events remains limited. For example the Belgian application Van Steenbergen et al., (2012) shows the performance of uncertainty spread using the same time period upon which the uncertainty matrix was constructed. This only shows the performance of the method on past uncertainty, not on the predictive uncertainty. The representativeness of the uncertainty bands for the forecasts remains unknown. There are opportunities using existing methods and techniques to assess and adjust the representative of a model ensemble or uncertainty bands for the predictive uncertainty, examples include (Abramowitz and Gupta, 2008; Madadgar et al., 2014). A research question that arises is: 'How can the representativeness of uncertainty bands around flood forecasts be assessed?'. A review is needed of the available metrics which assess representativeness of uncertainty spread highlighting their strengths, weaknesses and application.

The **fourth challenge** is that both methods struggle to represent uncertainty without **observed data**, the highest uncertainty is often related to ungauged catchments and the statistical method cannot be applied without data. The transferability of uncertainties from gauged to ungauged catchment has been explored (Bourgin et al., 2015), but more research is required. For the ensemble methods, in defining the sampling space for ensembles observed data is used. Both methods are dependent on observed data in testing the representativeness of the uncertainty spread. A research question that remains to be answered is: how can the representatives of uncertainty of an ungauged (sub) catchment be assessed? The Prediction in Ungauged Basins (PUB) initiative sought to answer this question and offered ways forwards which include the use and assimilation of satellite data (Hrachowitz *et al.*, 2013). The PUB work needs not only to be continued, but also requires an equivalent focussing on Forecasting in Ungauged Basins.

The **fifth challenge** is to generate uncertainty information which can be **used by *decision makers***. Decision makers work in a complex environment in which scientific information is often not prioritised above regulatory, institutional, political, resource and other constraints (Morss et al., 2005). Recent work by (Leskens et al., 2014) shows that the usefulness of flood forecasting model outputs to decision makers depends on the type and quality of the output and the flexibility and speed of the model. In order to generate useful uncertainty information for decision makers, they need to be incorporated into the research and development process. This cannot be done without long-term partnerships between scientist, product developers and different groups of decision makers (Meo et al., 2002) and (NRC, 2004) and effective communication (Bruen et al., 2010).

The research priority emerging from this discussion is the need to apply uncertainty quantification methods and theories more widely to catchments and flood events in a comparable way.

7. Conclusions

This review aims to provide practitioners with information to help match an uncertainty quantification method to their application. This is not straightforward owing to the variety of flood forecasting and warning systems. This paper has focussed on two specific applications of two well-known types of uncertainty quantification methods: statistical and ensemble. This research concludes that the statistical uncertainty quantification methods can answer the question: what is the probability of the forecasts being accurate, based on past performance? This method should be applied when the users require an estimation of the uncertainty of a flood forecast as probabilistic bands based on the historical uncertainty. The question that the ensemble methods can answer is: what is the spread of the forecast given the known lack of knowledge on the model structure, parameter and/or initial conditions of the catchment, river and/or atmosphere? The application of this approach would be targeted at a forecast system where uncertainty information from a specific source is required, for example, uncertainty from the meteorological forecast. Another application is where the forecaster needs uncertainty information at ungauged locations as well as gauged location in the catchment. The advantages of the statistical methods are mostly practical, related to the low computational requirements and resources needed and the fact that it can be bolted onto an existing system as a post process. The drawbacks are that the outputs are limited to locations with observed data and uncertainty cannot be split out into different types. The strength of the ensemble method lies in the fact that it treats the uncertainty from the source leading to more information on uncertainty which can be disaggregated, tracked through the cascade of models and through lead times, and is available in locations without observed data. The drawbacks are related to the computational power required to run an ensemble and the resources required for post processing, analysing and archiving the additional volumes of data. In conclusion, both methods are able to improve on a deterministic forecast and the choice will be a trade-off between the information required, available resources and the available data. Using available methods and theories from literature together with this overview guiding practitioners it should be possible to estimate uncertainty and produce a probabilistic forecast using any flood forecasting system.

8. Acknowledgments

This work forms part of the first author’s PhD research at HR Wallingford as an Affiliated Research Centre of the Open University. The Authors wish to thank Elisabeth Stephens and Antione Thiboult for their discussion and support during the early stages of this research.

9. References

Abramowitz, G., Gupta, H., 2008. Toward a model space and model independence metric. Geophys. Res. Lett. 35, L05705. https://doi.org/10.1029/2007GL032834

Beven, K., 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. 61, 1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven, K., 2012. Causal models as multiple working hypotheses about environmental processes. Comptes Rendus Geosci. 344, 77–88.

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897–5918.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 11–19.

Boucher, M.-A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012. Hydro-economic assessment of hydrological forecasting systems. J. Hydrol. 416, 133–144. http://dx.doi.org/10.1016/j.jhydrol.2011.11.042

Bourgin, F., Andréassian, V., Perrin, C., Oudin, L., 2015. Transferring global uncertainty estimates from gauged to ungauged catchments. Hydrol. Earth Syst. Sci. 19, 2535–2546. https://doi.org/10.5194/hess-19-2535-2015

Brochero, D., Gagné, C., Anctil, F., 2013. Evolutionary multiobjective optimization for selecting members of an ensemble streamflow forecasting model, in: Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation. ACM, pp. 1221–1228.

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

Cauwenberghs, K., Maeghe, K., 2007. Flood forecasting and early warning: state of play and needs in Flanders (Proceeds of EU Exchange Circle: “Flood Forecasting & Early Warning”).

Centre d’expertise hydrique du Québec, MDE, 2017. Hydrological and Hydraulic Forecasts [WWW Document]. Hydrol. Hydraul. Forecasts. URL https://www.cehq.gouv.qc.ca/prevision/index.asp

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Cloke, H.L., Pappenberger, F., 2008. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. Meteorol. Appl. 15, 181–197. https://doi.org/10.1002/met.58

Der Kiureghian, A., Ditlevsen, O., 2007. Special Workshop on Risk Acceptance and Risk Communication. Presented at the Special Workshop on Risk Acceptance and Risk Communication, Stanford University, Stanford University California.

Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrol. Earth Syst. Sci. 13, 913–921.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367. https://doi.org/10.1007/s10236-003-0036-9

FHR, F.H.R., 2017. waterinfo.be [WWW Document]. Water Info. URL http://www.waterinfo.be/default.aspx?path=NL/Thema/Overstroming_Actueel&KL=en (accessed 8.21.17).

Fortin, J., Moussa, R., Bocquillon, C., Villeneuve, J., 1995. Hydrotel, un modèle hydrologique distribué pouvant bénéficier des données fournies par la télédétection et les systèmes d’information géographique. J. Water Sci. 8, 97–124.

Fortin, V., Abaza, M., Anctil, F., Turcotte, R., 2014. Why should ensemble spread match the RMSE of the ensemble mean? J. Hydrometeorol. 15, 1708–1713.

Gotzinger, J., Bardossy, A., 2007. Generic error model for calibration and uncertainty estimation of hydrological models. Water Resouces Res. 44, W00B07. https://doi.org/10.1029/2007WR006691

Gouldby, B. P. and Samuels P. G. (2009) *Language of Risk: Project Definitions (Second edition)*, FLOODsite project report T32-04-01, available at http://www.floodsite.net/html/partner_area/project_docs/T32_04_01_FLOODsite_Language_of_Risk_D32_2_v5_2_P1.pdf (downloaded on 19 November 2018)

Gupta, H.V., Sorooshian, S., Hogue, S., Boyle, D.P., 2003. Advances in automatic calibration of watershed models. Calibration Watershed Models 9–28. https://doi.org/10.1029/WS006p0009

Gutmann, E., Pruitt, T., Clark, M.P., Brekke, L., Arnold, J.R., Raff, D.A., Rasmussen, R.M., 2014. An intercomparison of statistical downscaling methods used for water resource assessments in the United States. Water Resour. Res. 50, 7167–7186.

Hemri, S., Fundel, F., Zappa, M., 2013. Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. Water Resour. Res. 49, 6744–6755.

HEPEX, 2017. HEPEX - Implementation Plan for The Hydrological Ensemble Prediction Experiment [WWW Document]. Community Res. Pract. Adv. Hydrol. Ensemble Predict. URL www.hepex.org (accessed 8.9.16).

Heylen, J., 1997. De hydrologie van het IJzerbekken. Tijdschr. Waterprobl., Referaten van de internationale studiedag “Naar een integraal waterbeleid in het IJzerbekken” 16, 239–244.

Hrachowitz, M., Savenije, H.H.G., Bloschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB) - a review. Hydrol. Sci. J. 58, 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Juston, J.M., Kauffeldt, A., Quesada Montano, B., Seibert, J., Beven, K.J., Westerberg, I.K., 2013. Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling. Hydrol. Process. 27, 1117–1122. https://doi.org/10.1002/hyp.9625

Klein, B., Meissner, D., Kobialka, H.-U., Reggiani, P., 2016. Predictive Uncertainty Estimation of Hydrological Multi-Model Ensembles Using Pair-Couple Construction. Water 8. https://doi.org/10.3390/w8040125

Koenker, R., 2005. Quantile Regression, Econometric Society Monographs. Cambridge University Press.

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

Leedal, D., Neal, J., Beven, K., Young, P., Bates, P., 2010. Visualization approaches for communicating real-time flood forecasting level and inundation information. J. Flood Risk Manag. 3, 140–150. https://doi.org/10.1111/j.1753-318X.2010.01063.x

Leskens, J., Brugnach, M., Hoekstra, A.Y., Schuurmans, W., 2014. Why are decisions in flood disaster management so poorly supported by information from flood models? Environ. Model. Softw. 53, 53–61.

Li, H., Luo, L., Wood, E.F., Schaake, J., 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. J. Geophys. Res. 114, D04114. https://doi.org/10.1029/2008JD010969

Li, Y., Chen, J., Feng, L., 2013. Dealing with Uncertainty: A Survey of Theories and Practices. IEEE Trans. Knowl. Data Eng. 25, 2463–2482.

Li, Y., Grimaldi, S., Walker, J.P., Pauwels, V.R.N., 2016. Application of Remote Sensing Data to Constrain Operational Rainfall-Driven Flood forecasting: A Review. Remote Sens. 8, 456. https://doi.org/doi:10.3390/rs8060456

Liguori, S., Rico-Ramirez, M.A., 2014. A review of current approaches to radar-based quantitative precipitation forecasts. Int. J. River Basin Manag. 12, 391–402. https://doi.org/10.1080/15715124.2013.848872

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an intergrated data assimilation framework. Water Resouces Res. 43, W07401. https://doi.org/10.1029/2006WR005756

Madadgar, S., Moradkhani, H., Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrol. Process. 28, 104–122.

Madsen, H., Skotner, C., 2005. Adaptive state updating in real-time river flow forecasting - a combined filtering and error forecasting procedure. J. Hydrol. 308, 302–312. https://doi.org/10.1016/j.jhydrol.2004.10.030

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Theme\s sl, M., others, 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48.

Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continuous probability distributions. Manag. Sci. 22, 1087–1096.

Matte, S., Boucher, M.-A., Boucher, V., Fortier Filion, T.-C., 2017. Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-averse decision maker. Hydrol. Earth Syst. Sci. 2967–2986. https://doi.org/10.5194/hess-21-2967-2017

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26, 4078–4111. https://doi.org/10.1002/hyp.9384

Meo, M., Focht, W., Caneday, L., Lynch, R., Moreda, F., Pettus, B., Sankowski, E., Trachtenberg, Z., Vieux, B., Willett, K., 2002. Negotiating science and values with stakeholders in the Illinois River Basin. JAWRA J. Am. Water Resour. Assoc. 38, 541–554.

Montanari, A., 2007. What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology. Hydrol. Process. 841–845. https://doi.org/10.1002/hyp.6623

Morss, R.E., Wilhelmi, O.V., Downton, M.W., Gruntfest, E., 2005. Flood risk, uncertainty, and scientific information for decision making: lessons from an interdisciplinary project. Bull. Am. Meteorol. Soc. 86, 1593–1601.

NRC, N.R.C. (US) C. on W.R. for S., 2004. Where the weather meets the road: A research agenda for improving road weather services. National Academy Press.

Palmer, T.N., 2000. Predicting uncertainty in forecasts of weather and climate. Rep. Prog. Phys. 63, 71–116.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., de Roo, A.P.J., 2005. Cascading model uncertainty from medium range weather forecasts (10days_ through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrol. Earth Syst. Sci. 9, 381–393.

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., Krahe, P., 2011. The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast system. Atmospheric Res. 100, 150–167. https://doi.org/10.1016/j.atmosres.2010.11.016

Schoof, J., 2013. Statistical downscaling in climatology. Geogr. Compass 7, 249–265.

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. https://doi.org/10.5194/hess-1116-1171-2012

Smith, P., Pappenberger, F., Wetterhall, F., Thielen, J., Krzeminski, B., Salamon, P., Muraro, D., Kalas, M., Baugh, C., 2016. On the operational implementation of the European Flood Awareness System (EFAS) (Technical memorandum No. 778).

Stanski, H.R., Wilson, L.J., Burrows, W.R., 1989. Survey of common verification methods in meteorology.

Thiboult, A., Anctil, F., 2015a. Assessment of a multimodel ensemble against an operational hydrological forecasting system. Can. Water Resour. JournalRevue Can. Ressour. Hydr. 40, 272–284.

Thiboult, A., Anctil, F., 2015b. 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. https://doi.org/10.5194/hess-20-1809-2016

Todini, E., 2017. Flood Forecasting and Decision Making in the new Millennium. Where are We? Water Resour. Manag. 31, 3111–3129. https://doi.org/10.1007/s11269-017-1693-7

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

Van Steenbergen, N., Ronsyn, J., Willems, P., 2012. A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. Environ. Model. Softw. 33, 92–105. https://doi.org/10.1016/j.envsoft.2012.01.013

Van Steenbergen, N., Willems, P., 2015. Uncertainty decomposition and reduction in river flood forecasting: Belgian case study. J. Flood Risk Manag. 8, 263–275. https://doi.org/10.1111/jfr3.12093

Verkade, J., Brown, J., Reggiani, P., Weerts, A., 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. J. Hydrol. 501, 73–91.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. Shuffled Complex Evolution Metropolis Algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resouces Res. 39, 1201. https://doi.org/doi:10.1029/2002WR001642

Wani, O., Beckers, J.V., Weerts, A.H., Solomatine, D.P., 2017. Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. Hydrol. Earth Syst. Sci. 21, 4021.

Weerts, A.H., Winsemius, H.C., Verkade, J.., 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.

Zappa, M., Beven, K.J., Bruen, M., Cofino, A.S., Kok, K., Martin, E., Nurmi, P., Orfila, B., Roulin, E., Schroter, K., Seed, A., Szturc, J., 2010. Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. Atmospheric Sci. Lett. 11, 83–91. https://doi.org/10.1002/asl.248

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