Hydrol. Earth Syst. Sci. Discuss., 11, 10179–10233, 2014
www.hydrol-earth-syst-sci-discuss.net/11/10179/2014/
doi:10.5194/hessd-11-10179-2014

Hydrology and Earth System Sciences
Open Access
Discussions

Estimation of predictive hydrologic uncertainty using quantile regression and UNEEC methods and their comparison on contrasting catchments

N. Dogulu[1,*], **P. López López**[1,2,**,***], **D. P. Solomatine**[1,3], **A. H. Weerts**[2,4], and **D. L. Shrestha**[5]

[1]UNESCO-IHE Institute for Water Education, Delft, the Netherlands
[2]Deltares, Delft, the Netherlands
[3]Water Resources Section, Delft University of Technology, Delft, the Netherlands
[4]Wageningen University and Research Centre, Hydrology and Quantitative Water Management Group, Wageningen, the Netherlands
[5]CSIRO Land and Water, Highett, Victoria, Australia
[*]now at: Middle East Technical University, Ankara, Turkey
[**]now at: Utrecht University, Utrecht, the Netherlands
[***]now at: Deltares, Delft, the Netherlands

Received: 14 July 2014 – Accepted: 3 August 2014 – Published: 10 September 2014

Correspondence to: N. Dogulu (ndogulu@metu.edu.tr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD
11, 10179–10233, 2014

Abstract

In operational hydrology, estimation of predictive uncertainty of hydrological models used for flood modelling is essential for risk based decision making for flood warning and emergency management. In the literature, there exists a variety of methods analyzing and predicting uncertainty. However, case studies comparing performance of these methods, most particularly predictive uncertainty methods, are limited. This paper focuses on two predictive uncertainty methods that differ in their methodological complexity: quantile regression (QR) and UNcertainty Estimation based on local Errors and Clustering (UNEEC), aiming at identifying possible advantages and disadvantages of these methods (both estimating residual uncertainty) based on their comparative performance. We test these two methods on several catchments (from UK) that vary in its hydrological characteristics and models. Special attention is given to the errors for high flow/water level conditions. Furthermore, normality of model residuals is discussed in view of clustering approach employed within the framework of UNEEC method. It is found that basin lag time and forecast lead time have great impact on quantification of uncertainty (in the form of two quantiles) and achievement of normality in model residuals' distribution. In general, uncertainty analysis results from different case studies indicate that both methods give similar results. However, it is also shown that UNEEC method provides better performance than QR for small catchments with changing hydrological dynamics, i.e. rapid response catchments. We recommend that more case studies of catchments from regions of distinct hydrologic behaviour, with diverse climatic conditions, and having various hydrological features be tested.

1 Introduction

Importance of accounting for uncertainty in hydrological models used in flood early warning systems is widely recognised (e.g. Krzysztofowicz, 2001; Pappenberger and Beven, 2006). Such an uncertainty in the model prediction stems mainly from the four important sources: perceptual model uncertainty, data uncertainty, parameter estimation uncertainty, and model structural uncertainty (e.g. Solomatine and Wagener, 2011). Estimation of *predictive uncertainty* (Coccia and Todini, 2011) of hydrological models used for flood modeling enable hydrologists and managers to achieve better risk based decision making and thus has the potential to increase the reliability and credibility of flood warning. Therefore, the necessity of estimating predictive uncertainty of rainfall–runoff models is broadly acknowledged in operational hydrology, and the management of uncertainty in hydrologic predictions has emerged as a major focus of interest in both research and operational modelling (Wagener and Gupta, 2005; Liu and Gupta, 2007; Montanari, 2007; Todini, 2008). In this respect comparing different methods, which are often developed and tested in isolation, receives attention of researchers, e.g. as suggested within the HEPEX framework (see van Andel et al., 2013).

While the discussions on the necessity of evaluating the contribution of various sources of errors to the overall model uncertainty are going for a long time (see, e.g. Gupta et al., 2005; Brown and Heuvelink, 2005; Liu and Gupta, 2007), there have been also attempts to estimate the *residual uncertainty*. By residual uncertainty, we understand the remaining model uncertainty assuming that other sources were accounted for (for example by calibrating the parameters), or not considered (all other sources like inaccurate rating curve, inputs, etc.) (Solomatine and Shrestha, 2009). We recognize that there are many sources of uncertainty leading to uncertainty in the model output (their influence is typically explored by running Monte Carlo experiments). However in this paper we consider the uncertainty of model outputs, assuming that parameters, inputs and the data used for model calibration are known (so we don't consider their uncertainty explicitly). Within this context, a (residual) model error is seen as a manifestation of the (residual) model uncertainty.

To analyze and capture residual uncertainty, statistical methods are often used. The prediction bounds are estimated by either purely statistical methods, e.g. meta-Gaussian approach (Montanari and Brath, 2004; Todini, 2008; Regianni and Weerts, 2008; Regianni et al., 2009), quantile regression (QR; Solomatine and Shrestha; 2009; Weerts et al., 2011), regression analysis on absolute errors (DUMBRAE; Pianosi and Raso, 2012); or machine learning techniques (UNEEC; see, e.g. Shrestha and Solomatine, 2006; Solomatine and Shrestha; 2009), and wavelet analysis (Bogner and Pappenberger, 2011). In this paper we consider two methods that differ in their methodological complexity: quantile regression (QR) and UNcertainty Estimation based on local Errors and Clustering (UNEEC).

Quantile regression (Koenker and Basset, 1978; Koenker and Hallock, 2001; Koenker, 2005) is a statistical regression technique that models the relationship between one or more predictors (inputs) and the predictand (response variable). In QR, a regression model is developed for selected quantiles of the conditional distribution of the response variable (discharge or water level in the present research study). This methodology allows for examining the entire distribution of the variable of interest rather than a single measure of the central tendency of its distribution (Koenker, 2005). QR models have been used in a broad range of applications, such as economics and financial market analysis (Kudryavtsev, 2009; Taylor, 2007), agriculture (Barnwal and Kotani, 2013), meteorology (Bremnes, 2004; Friederichs and Hense, 2007; Cannon, 2011), wind forecasting (Nielsen et al., 2006; Møller et al., 2008), the prediction of ozone concentrations (Baur et al., 2004; Munir et al., 2012), etc. In hydrological modelling the QR method has been applied as an uncertainty post-processing technique in previous research studies with different configurations. The configurations differ mainly in two aspects: treatment of quantiles crossing problem and the quantiles derivation in normal space using the Normal Quantile Transformation (NQT). Solomatine and Shrestha (2009) make use of the classical QR approach, without considering quantiles crossing and NQT. Weerts et al. (2011), Verkade and Werner (2011), and Roscoe et al. (2012) apply QR to various deterministic hydrologic forecasts. QR configuration investigated in these studies uses the water level or discharge forecasts as predictors to estimate the distribution quantiles of the model error. It includes a transformation into normal space using the NQT and the quantile crossing problem is addressed imposing a fixed

HESSD
11, 10179–10233, 2014

distribution of the predictand in the crossing domain. Singh et al. (2013) make use of a similar configuration differentiating two cases based on the similarities in information content between calibration and validation data periods. However, López López et al. (2014) apply QR to predict the quantiles of the environmental variables itself (water level) rather than the quantiles of the model error, and the four different configurations of QR are compared and extensively verified.

UNEEC was introduced in 2006 (Shrestha and Solomatine, 2006; Shrestha et al., 2006). The method builds a regression model to estimate the quantiles of the error distribution; however it is not an autoregressive model (as in QR). UNEEC employs more complicated machine learning approaches and is based on the recognition that residual uncertainty depends on a number of variables characterising the state of the modelled system. Another notable characteristic of UNEEC is the local modelling of errors (through clustering) so that particularities of different hydrometeorological conditions, i.e. heterogeneities inherent in rainfall–runoff process, are represented through different error *pdf*s. Shrestha and Solomatine (2006) tested the UNEEC method on Sieve catchment in Italy based on the estimates of lower and upper prediction limits corresponding to 90 % confidence level. The method was also applied to a different catchment (Brue, in UK; HBV model) and its performance was compared with GLUE (Beven and Binley, 1992) and meta-Gaussian approach (Montanari and Brath, 2004). It was reported that the uncertainty estimates obtained by UNEEC were in fact more acceptable and interpretable than those obtained by the other methods. UNEEC was further extended to estimate several quantiles (thus approximating full *pdf* of the error distribution) and applied to Bagmati catchment in Nepal (Solomatine and Shrestha, 2009), and it was compared to several other methods including QR. It was found that UNEEC method generated consistent and interpretable results which are more accurate and reliable than QR. Pianosi et al. (2010) extended UNEEC so as to include parametric uncertainty (UNEEC-P), however local features of uncertainty were not considered. Nasseri et al. (2014) compared UNEEC with methods which are mainly based on the fuzzy extension principle: IMFEP (Incremental Modified Fuzzy Extension Principle)

HESSD
11, 10179–10233, 2014

and MFEP (Modified Fuzzy Extension Principle). It was seen that the methods provided similar performance on the two monthly water balance models for the two basins in Iran and France.

Solomatine and Shrestha (2009) presented their initial experiments to compare QR and UNEEC on one case study, and Weerts et al. (2011) discussed the experience with QR on another one. In this paper we go further and test the newer variants of these methods on several contrasting catchments that cover a wide range of climatic conditions and hydrological characteristics. The motivation here is to identify possible advantages and disadvantages of using QR and UNEEC methods based on their comparative performance, especially during flooding conditions (i.e. for the data cluster associated with high flow/water level conditions). The knowledge gaps regarding the use of the methods with different parameterizations are addressed. For example, we now incorporate in UNEEC the autoregressive component by considering past error values (in addition to discharge and effective rainfall) in one case study, and model outputs for the state variables soil moisture deficit (SMD) and groundwater level (GW) are used as predictors (in addition to water level) in another case study. In QR, the linear regression model was established to predict the quantiles of observed water levels conditioned on simulated/forecasted water levels. Furthermore, we present results of statistical analysis of error time series to better understand (hydrological) models' quality in relation to its effect on uncertainty analysis results, and to discuss the assumption of normality in the model residuals, particularly in view of the clustering approach employed within the framework of UNEEC method. We apply methods to estimate predictive uncertainty in Brue catchment (southwest UK) and Upper Severn catchments – Yeaton, Llanyblodwel, and Llanerfyl (Midlands, UK).

The remainder of the paper is structured as follows. The next section describes the residual uncertainty analysis methods (QR and UNEEC) and the validation measures used. Section 3 describes the studied catchments and the conducted experiments. The results for error and uncertainty analyses are presented and discussed in Sect. 4. In

Sect. 5 the main conclusions from the study and recommendations for future work are presented.

2 Methodology

2.1 Uncertainty analysis methods

As in Solomatine and Shrestha (2009) and Weerts et al. (2011), we consider a deterministic (hydrological) model M of a catchment predicting a system output variable $\hat{y}$ given the input data matrix $\mathbf{x}$, and the vector of model parameters $\boldsymbol{\theta}$. There are various sources of error associated with the model output (e.g. discharge), so the system response (i.e. actual discharge) can be expressed as:

$$\boldsymbol{y}_{t+\mathrm{LT}} = \hat{\boldsymbol{y}} + \boldsymbol{e} = M(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{e}$ is the total residual error (in the remainder of the text, the terms "model error" and "model residual" is used interchangeably to refer to $\boldsymbol{e}$). The model M can be used in two modes depending on the relation between the lead time (LT: the duration between time of forecast and time for which the forecast is made) of interest and the model time step (Δt):

$$\begin{cases} \text{simulation mode,} & \mathrm{LT} = 1 \cdot \Delta t \\ \text{forecasting mode,} & \mathrm{LT} > 1 \cdot \Delta t \end{cases} \quad (2)$$

Given the model structure M, and the parameter set $\boldsymbol{Q}$, the uncertainty analysis methods used in this study, namely QR and UNEEC, estimate the residual uncertainty of a calibrated hydrological model whose parameters and inputs are assumed to be known

HESSD
11, 10179–10233, 2014

exactly. In this setup the different sources of uncertainty are not distinguished explicitly. In both methods, the uncertainty model U predicts the quantile value q^{τ} and is calibrated for different quantiles (τ), and for various lead times (LT) separately:

$$\boldsymbol{q}^{\tau}_{t+\mathrm{LT}} = U(\mathbf{I}, \boldsymbol{\lambda}) \tag{3}$$

where $\mathbf{I}$ is the input data matrix, and $\boldsymbol{\lambda}$ is the vector of model parameters. In a simplest case when number of quantiles is 2, they form the confidence level (e.g. 90 %) and the corresponding confidence interval, CI. The quantiles computed in this study are $\tau = 0.05$, 0.25, 0.75, and 0.95 allowing for forming the 50 and 90 % confidence intervals.

2.1.1 Quantile regression

As mentioned, several QR configurations have been previously investigated for estimating the residual uncertainty. Last research by López López et al. (2014) compares and verifies four alternative configurations of QR for several catchments at the Upper Severn River. The comparative analysis includes different experiments on the derivation of regression quantiles in original and in normal space using NQT, a piecewise linear configuration considering independent predictand domains and avoiding the quantiles crossing problem with a relatively recent technique (Bondell et al., 2010). Results show similar performance with all configurations in terms of reliability, sharpness and resolution. Due to this, the variant called "QR1: non-crossing Quantile Regresssion" was applied in the present study. QR1 estimates the quantiles of the distribution of water level or discharge in the original domain, without any initial transformation and avoids the quantiles crossing problem with the methodology proposed by Bondell et al. (2010). A brief description of the QR configuration used in the present work is given below (for details the reader is referred e.g. to López López et al., 2014).

For every quantile τ, we assume a linear relationship between the forecasted (or predicted) value, $\hat{s}$, and the real observed value, s,

$$s = a_{\tau}\hat{s} + b_{\tau} \tag{4}$$

where a_τ and b_τ are the parameters of linear regression. By minimising the sum of residuals, one can find the parameters a_τ and b_τ:

$$\min \sum_{j=1}^{J} \rho_\tau(s_j - (a_\tau \hat{s}_j + b_\tau)) \tag{5}$$

where s_j and $\hat{s}_j$ are the jth paired samples from a total of J samples and ρ_τ is the quantile regression function for the quantile τ:

$$\rho_\tau(\varepsilon_j) = \begin{cases} (\tau-1)\cdot\varepsilon_j, & \varepsilon_j \leq 0 \\ \tau\cdot\varepsilon_j, & \varepsilon_j \geq 0 \end{cases} \tag{6}$$

Equation (6) is applied for the error (ε_j), which is defined as the difference between the observation (s_j) and the linear QR estimate ($a_\tau \hat{s}_j + b_\tau$) for the selected quantile τ.

Figure 1 illustrates the estimation of a selection of quantiles, including 0.95, 0.75, 0.25 and 0.05 quantiles. To obtain the QR function for a specific quantile, e.g. $\tau = 0.05$, Eqs. (5) and (6) are applied as follows:

$$\rho_{0.05}(\varepsilon_j) = \begin{cases} -0.95\cdot\varepsilon_j, & \varepsilon_j \leq 0 \\ 0.05\cdot\varepsilon_j, & \varepsilon_j \geq 0 \end{cases} \tag{7}$$

In case of an ideal model, the 95 % of observed-forecasted pairs would be located above $\tau = 0.05$ quantile linear regression line, and 5 % would remain below it. Considering the two observed-forecasted pairs of the total of J samples, $j = 1$ and $j = 2$, their corresponding errors, ε_1 and ε_2, are:

$$\begin{aligned} \varepsilon_1 &= s_1 - (a_{0.05}\hat{s}_1 + b_{0.05}) < 0 \\ \varepsilon_2 &= s_2 - (a_{0.05}\hat{s}_2 + b_{0.05}) > 0 \end{aligned} \tag{8}$$

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Introducing both values in Eq. (5), QR allows for solving the minimization problem calculating the regression parameters $a_{0.05}$ and $b_{0.05}$ for this particular quantile $\tau = 0.05$:

$$\min(-0.95 \cdot \varepsilon_1 + 0.05 \cdot \varepsilon_2 + \ldots + \rho_{0.05}(\varepsilon_J)) \quad (9)$$

The procedure explained here can be extended for any quantile, τ.

2.1.2 UNEEC

In UNEEC, a machine learning model, e.g. an artificial neural network, model is built to predict uncertainty associated to model outputs for the future inputs to the hydrological model. The steps involved in UNEEC are summarized below:

- Identify the set predictor variables (e.g. the lagged rainfall data, soil moisture, flow, etc.) that describe the flow process based on their effect on the model error. These predictors can be selected using Average Mutual Information (AMI) and correlation analysis. Using AMI brings the advantage of detection of nonlinear relationships (Battiti, 1994).

- Employ the fuzzy c-means method to derive the fuzzy clusters in the data where predictors are the same or different predictors used in machine learning model, and the model error is the output attribute (Fig. 2). The use of fuzzy c-means allows for reflection of the smooth nature of variability in hydrological variables and provides a gradual transition between local error models identified by clusters formed. The optimal number of clusters can be determined using the existing methods, e.g. Xie and Benie (1991), Halkidi et al. (2001), Nasseri and Zahraie (2011).

- For each cluster c, calculate the quantiles, q_c^τ, of the empirical distribution of the model error.

HESSD
11, 10179–10233, 2014

N. Dogulu et al.

- For each data vector, calculate the "global" estimate of the quantile q^τ using the calculated quantiles q_c^τ. This is done by weighting the cluster quantile by the corresponding degree of membership of the given data vector to this cluster. Calculated q^τ values for each quantile τ are used as outputs for the uncertainty model U.
- Train a machine learning model (U) (e.g. ANN) using the set of predictors as inputs, and the data prepared at the previous step as the output. U will be able to predict the quantile value q^τ for the new input vectors.

2.2 Validation methods

In this study we use several statistical measures of uncertainty to evaluate and to some extent to compare performances of QR and UNEEC. These are, namely, mean prediction interval (MPI; Shrestha and Solomatine, 2006), prediction interval coverage probability (PICP; Shrestha and Solomatine, 2006), average relative interval length (ARIL; Jin et al., 2010), and normalized uncertainty efficiency (NUE; Nasseri and Zahraie, 2011). MPI and PICP have been widely used in the literature.

MPI computes the average width of uncertainty band (or prediction interval), i.e. the distance between upper and lower prediction limits ($\mathrm{PL}_t^{\mathrm{upper}}$ and $\mathrm{PL}_t^{\mathrm{lower}}$, respectively):

$$\mathrm{MPI} = \frac{1}{n}\sum_{t=1}^{n}\left(\mathrm{PL}_t^{\mathrm{upper}} - \mathrm{PL}_t^{\mathrm{lower}}\right) \tag{10}$$

MPI = 0 means there is no uncertainty at all. MPI is rather simple indicator giving an idea about the distribution sharpness.

PICP, on the other hand, is a more informative uncertainty indicator measuring the probability that the observed values (y_t) lie within the estimated prediction limits computed for a significance level of $1-\alpha$ (e.g. 90 %):

$$\mathrm{PICP} = \frac{1}{n}\sum_{t=1}^{n} C \quad \text{where } C = \begin{cases} 1, & \mathrm{PL}_t^{\mathrm{lower}} \le y_t \le \mathrm{PL}_t^{\mathrm{upper}} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

HESSD
11, 10179–10233, 2014

Ideally, PICP value should be equal or close to specified confidence level.

ARIL is similar to MPI and considers average width of uncertainty bounds in relation to the observed value:

$$\mathrm{ARIL} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left(\mathrm{PL}_t^{\mathrm{upper}} - \mathrm{PL}_t^{\mathrm{lower}}\right)}{y_t} \quad (12)$$

Having the observed value in denominator accounts for the fact that uncertainty (and MPI) is usually higher for higher values of flow and thus has a "normalization" effect. A problem with ARIL is that if the flow is zero or close to zero, ARIL will be infinity or very high.

There is no single objective measure of the quality of an uncertainty prediction method (since the "actual" uncertainty of the model is not known). Closer PICP is to the confidence level higher the trust in a particular uncertainty prediction method should be. In principle, a reliable method should lead to reasonably low values of MPI (and ARIL).

A possibility to combine PICP and ARIL is to use the NUE indicator:

$$\mathrm{NUE} = \frac{\mathrm{PICP}}{w \times \mathrm{ARIL}} \quad (13)$$

(in this study, the value of scale factor w is taken as 1) Nasseri and Zahraie (2011) recommend that methods with the higher NUE should be preferred over those with the lower NUE, however we do not think this is a universally applicable recommendation: if for two methods PICP is equal and close to the confidence interval (90 %) and ARIL for one method is higher (which is not good), then NUE for this method will be actually lower.

We would like to stress again that none of the presented measures allow for accurate comparison between different methods of uncertainty prediction (since the actual model uncertainty is never known), and should be therefore seen only as indirect indicators of methods' performance. These average measures should be used together with the uncertainty bound plots which visual analysis reveals more information on the capacity of different uncertainty prediction methods during particular periods.

3 Application

3.1 Case studies

3.1.1 Brue catchment

Located in the southwest of England, the Brue River catchment has a history of severe flooding. Draining an area of 135 km^2 to its river gauging station at Lovington (Fig. 3a), the catchment is predominantly rural and of modest relief and gives rise to a responsive flow regime due to its soil properties. The major land use is pasture on clay soil. The mean annual rainfall in the catchment is 867 mm and mean river flow is 1.92 m^3 s^{-1} (basin average, 1961–1990) (Table 1). This catchment has been extensively used for research on weather radar, quantitative precipitation forecasting and rainfall–runoff modelling, as it has been facilitated with a dense rain gauge network (see, e.g. Moore et al., 2000; Bell and Moore, 2000).

The flow in Brue River was simulated by HBV-96 model (Lindström et al., 1997), which is an update version of the HBV rainfall–runoff model (Bergström, 1976). This lumped conceptual hydrological model consists of subroutines for snow accumulation and melt (excluded for Brue), soil moisture accounting procedure, routines for runoff generation, and a simple routing procedure (Fig. 3b). The input data used are hourly observations of precipitation (basin average), air temperature, and potential evapotranspiration (estimated by modified Penmann method) computed from the 15 min data. Model time step is one hour (Δt = 1 h). The model is calibrated automatically using adaptive cluster covering algorithm (ACCO) (Solomatine et al., 1999). The data sets used for calibrating and validating the HBV-96 model are based on Shrestha and

Solomatine (2008). It should be mentioned that the discharge data on calibration has many peaks which are higher in magnitude compared to those in the validation data.

The uncertainty analyses conducted for Brue catchment are based on one-step-ahead flow estimates, i.e. LT = 1 h (simulation mode). Effective rainfall (rainfall minus evapotranspiration) values were used instead of using rainfall data directly.

3.1.2 Upper Severn catchments

Flowing from Cambrian Mountains (610 m) in Wales, the River Severn is the longest river in Britain (about 354 km). It forms the border between England and Wales and flows into the Bristol Channel. The river drains an area of approximately 10 500 km^2 above the monitoring station at Upton on Severn. Mean annual precipitation ranges from approximately 2500 mm in the west to less than 700 mm in the south (EA, 2009). The Upper Severn includes rock formations classified as non-aquifers as well as loamy soils characterised by their high water retention capacity (for more detailed description of the Upper Severn, see Hill and Neal, 1997). Flooding is a major problem at the downstream due to excessive rainfall at the upstream (the Welsh hills), early 2014 floods being the most recent significant floods that occurred.

In this work, the three sub-catchments of Upper Severn River are analyzed: Yeaton, Llanyblodwel, and Llanerfyl (Fig. 4). The area, elevation, mean flow, mean annual rainfall and basin lag time (time of concentration) information of the catchments are presented in Table 1. Yeaton catchment is located at a lower elevation and over a flat area compared to Llanerfyl and Llanyblodwel. This catchment has also the longest basin lag time. The smallest catchment in terms of drainage area is Llanerfyl, which also has the shortest basin lag time (approx. 3–5 h) leading to flash floods, so that the predictive uncertainty information on flood forecast for this catchment has especially high importance.

In Midlands Flood Forecasting System (MFSS; a Delft-FEWS forecast production system as described in Werner et al., 2013), the Upper Severn catchment is represented by a combination of numerical models for: rainfall–runoff modelling (MCRM;

HESSD
11, 10179–10233, 2014

Bailey and Dobson, 1981), hydrological routing (DODO; Wallingford, 1994), hydrodynamic routing (ISIS; Wallingford, 1997), and error correction (ARMA). The input data used within MFSS includes (a) Real Time Spatial data (observed water level and raingauge data as well as air temperature and catchment average rainfall); (b) Radar Actuals, (c) Radar Forecasts, and (d) Numerical Weather Prediction data (all provided by the UK Meteorological Office). The data available was split into two parts for calibration (7 March 2007 08:00–7 March 2010 08:00) and validation (7 March 2010 20:00–7 March 2013 08:00), preserving similar statistical properties in both data sets.

The forecasting system issues two forecasts per day (08:00 and 20:00 UTC) with a time horizon of two days. First, the estimates of internal states are obtained running the models (which are forced with observed precipitation, evapotranspiration and temperature) in historical mode over the previous period. The state variables for the (hydrological) model are soil moisture deficit (SMD, the amount of water required to bring the current soil moisture content to field capacity in the root zone), groundwater level (GW), snow water equivalent (SWE), and snow density (SD). Using a standalone version of MFSS, the system (forced by the forecasted precipitation) is then run forward with a time step of 1 h.

It is important to note that this case study, unlike Brue catchment, includes errors in the meteorological forecast and the back transformation of discharge to water level –via rating curve – in a lumped manner. Therefore, the effects of *rating curve uncertainty* (Di Baldassarre and Montanari, 2009; Sikorska et al., 2013; Coxon et al., 2014; Mukolwe et al., 2014) and *precipitation forecast uncertainty* (Kobold and Sušelj, 2005; Shrestha et al., 2013) are accommodated as well.

The uncertainty analysis is aimed at estimating predictive uncertainty for the forecast time series ($\Delta t = 12$ h) corresponding to the lead time of interest. In this study, we consider the lead times LT = 1, 3, 6, 12, and 24 h only.

3.2 Experimental setup

The experiments conducted in this study focus mainly on four aspects: (a) assessment of hydrological model quality in support of understanding its effect on uncertainty analysis results, (b) investigation of normality of model residuals, particularly in view of clustering approach employed within the framework of UNEEC method, (c) proper setup of the QR and UNEEC algorithms, and (d) evaluation of uncertainty analysis results from the both methods based on their comparative performance.

The aspects related to (a) and (b) are presented for Brue catchment in Sect. 4.1.1 and for Upper Severn catchments in Sect. 4.1.2. In order to assess hydrological model quality, we analyze error time series statistically. Normality of model residuals is investigated through probability plots of a *normal* distribution and a *t location-scale* distribution, which *pdf* is given by Eqs. (14) and (15), respectively.

$$f(x) = \left(1/\sigma\sqrt{2\pi}\right) e^{-(x-\mu)^2/2\sigma^2} \tag{14}$$

$$f(x) = \frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\sigma\sqrt{2\pi}\,\Gamma\left(\nu+\frac{1}{2}\right)}\left[\frac{\nu+\left(x-\frac{\mu}{\sigma}\right)^2}{\nu}\right]^{-\left(\nu+\frac{1}{2}\right)} \tag{15}$$

where μ: location parameter (mean), σ: scale parameter (std. deviation), ν: shape parameter (i.e., the number of degrees of freedom), and Γ: gamma function. The *t* location-scale distribution is like the normal distribution by definition; however, it has heavier tails making it more prone to outliers. Within this study outliers refer to very high model residuals occurring during extreme precipitation and flow events. In case of normality of data its analysis becomes much simpler, however often this is not the case.

Residual uncertainty varies in time and with the changing hydrometeorological situation, and in this paper we investigate residuals distribution for different hydrometeorological conditions represented by clusters found within the UNEEC method.

HESSD
11, 10179–10233, 2014

Sections 4.2.1 and 4.2.2 describe the aspects related to (c) and (d) for Brue catchment and Upper Severn catchments, respectively. The model setups for QR and UNEEC are identified based on their methodological description and catchment specific conditions such as data availability. Selection of best model setup for UNEEC includes AMI and correlation analysis, and determination of number of clusters.

In the Brue catchment case study, for QR, a linear regression model estimating quantile τ of observed discharge conditioned on simulated discharge are obtained. UNEEC method builds more complex data-driven models estimating quantiles of error using more variables and considers local features of error through clustering. The methodological details on application of UNEEC method on Brue catchment is explained below.

Shrestha and Solomatine (2008) tested UNEEC method on Brue catchment to assess predictive uncertainty of the one-step-ahead flow estimates. The probable predictors of model error identified using AMI and correlation analysis were only lagged discharge ($\boldsymbol{Q}_{t-1}, \boldsymbol{Q}_{t-2}, \boldsymbol{Q}_{t-3}$) and effective rainfall ($\boldsymbol{RE}_{t-8}$, $\boldsymbol{RE}_{t-9}$, $\boldsymbol{RE}_{t-10}$) values. In this study, however, we try a different parameterization of the method. In addition to the mentioned variables, we consider the most recent two past error values ($\boldsymbol{e}_{t-1}, \boldsymbol{e}_{t-2}$) as predictors to incorporate the autoregressive features. It was seen that this configuration resulted in decreased MPI values (< 5 %) during both training and test periods. In accordance with the previous study the number of clusters used is 5. For the quantiles of interest, different M5 model tree (Quinlan, 1992) models are built. A model tree is a hierarchical (i.e. tree like) modular model which can be considered analogous to a piecewise linear function. At non-terminal nodes there are rules that progressively split data into subsets, and finally there are linear regression equations at the leaves of the tree built on the data subset that reached this particular leaf. Model trees can be easily used for tasks with very high dimensionality as it learns efficiently.

In the case studies from Upper Severn catchment, for QR, a linear regression model estimating quantile τ of observed water level conditioned on forecasted water levels is obtained. For UNEEC, different M5 model tree models are built. In the UNEEC method, the first step is to identify the exact structure of the model error predictor using AMI and correlation analysis. In these case studies a variety of predictors are considered for the model, e.g. observed and modelled water level, forecasted precipitation, and state variables (***GW***, ***SMD***, ***SWE***, ***SD***). Although the benefits of using the observed (and/or modelled) soil moisture and groundwater level information for modelling rainfall–runoff processes and predicting runoff is well known in the literature (Aubert et al., 2003; Lee and Seo, 2011; Tayfur et al., 2014), we cannot cite any studies exploring the possible advantages of using such information for improving predictive capabilities of uncertainty analysis methods. Therefore, the dependence of model residuals on variables expressing internal state of the catchments is also analyzed.

Among the state variables, the most significant correlation with the model error was of ***GW*** and ***SMD***. While ***GW*** was found to be positively correlated with model residuals (i.e. as ***GW*** increases, error increases too), ***SMD*** and model error had a negative correlation. High groundwater levels are associated with more precipitation. High soil moisture deficit, on the other hand, indicates that there has been no excessive precipitation and the soil is not filled up with infiltrated water. High soil moisture deficit might also occur due to increasing evaporation rates causing soil to dry up. However, considering climate of Upper Severn region, low soil moisture is more likely attributed to higher rainfall rates. Eventually, it was decided that the most recent precipitation ($\boldsymbol{P}_{t-1}$), observed water level ($\boldsymbol{H}_{\text{obs},t-1}$), error ($\boldsymbol{e}_{t-1}$), and state variables GW and SMD shall be considered as predictors. It should be noted that subscript $t-1$ stands for $t-12$ h in reality as the data sets analyzed has a time step of 12 h (see Sect. 3.1.2). For the sake simplicity, we prefer to use the subscript notation $t-1$.

A number of experiments with UNEEC have been conducted in order to be able to select the best model setup among the variables ***GW***, ***SMD***, $\boldsymbol{H}_{\text{obs},t-1}, \boldsymbol{P}_{t-1}, \boldsymbol{e}_{t-1}$. From both calibration and validation results it was seen that there was only negligible changes (and mostly no change) in terms of MPI and PICP when $\boldsymbol{P}_{t-1}$ and $\boldsymbol{e}_{t-1}$ were included. This is indeed a pertinent finding ensuring that no forecast error in precipitation is considered as an additional uncertainty source within the uncertainty model

U. Noting also that correlation between precipitation and model residuals was very much less than that of between observed water level and model residuals, P_{t-1} is not considered as an input vector for the model U. Unlike in Brue catchment, we did not include past model residuals ($\boldsymbol{e}_t$) as an input data vector as its possible effects might be highly misleading especially for longer lead times. Consequently, only the variables $\boldsymbol{GW}$, $\boldsymbol{SMD}$, $\boldsymbol{H}_{\text{obs},t-1}$ are selected as final predictors. The significance of $\boldsymbol{GW}$ variable is substantial in that inclusion of this variable provides more explainable results in terms of MPI and PICP. As such, the use of $\boldsymbol{GW}$ variable (together with $\boldsymbol{SMD}$) can be considered as a proxy for using rainfall information, thus is highly necessary.

The fuzzy c-means method was used with 6 clusters where fuzzy exponential coefficient was set to 2. M5 model tree was used as the machine learning model. Main reasons for using this technique are its accuracy, transparency (analytical expressions for models are obtained explicitly) and speed in training. The decision on optimal number of clusters was based on computation of Partition Index (SC), Separation Index (S) and Xie and Beni Index (XB) (Bensaid et al., 1996; Xie and Beni, 1991), and observing sensitivity of PICP and MPI.

Within the variables considered in clustering, $\boldsymbol{GW}$ is the most influential one. Fig. 5 shows fuzzy clustering of $\boldsymbol{GW}$, $\boldsymbol{SMD}$, and $\boldsymbol{H}_{\text{obs},t-1}$ data for Llanyblodwel catchment (lead time = 6 h). Also on the same figure is the plot of model residuals against $\boldsymbol{GW}$ where one can observe heteroscedasticity of model residuals with respect to $\boldsymbol{GW}$. As can be easily seen, while cluster 2 is associated to very high groundwater levels, cluster 4 can be attributed to low groundwater level conditions, which might occur due to low water levels in the river and/or high soil moisture deficit. Looking at groundwater level time series in Fig. 5, one can notice that the change in $\boldsymbol{GW}$ is approximately 60 m in the first three months period of calibration data (from 0 to time step 200). Such amount of change is too big for a process which is known to be considerably slower, e.g. as compared to river flow process. This can be explained by the fact that conceptual models are inaccurate and cannot be expected to reproduce all the complex physics of nature (groundwater being one of the most complex parts). There is also probably

HESSD
11, 10179–10233, 2014

a "compensatory" effect of the model, when one part of the model (groundwater) does something non-physical to try to end up (mathematically) with the reasonable values of output (flow).

Uncertainty analysis results from both methods are evaluated and compared employing the validation measures explained in Sect. 2.2.

4 Results and discussion

This part focuses on the statistical error analysis (Sect. 4.1) and compares uncertainty analysis results obtained from QR and UNEEC (Sect. 4.2).

4.1 Statistical error analysis

4.1.1 Brue catchment

Observed discharge plotted against simulated discharge during calibration period can be seen in Fig. 6a, whereas Fig. 6b shows how model residuals change with the observed discharge. As expected, model error increases with increasing discharge values. Although the model residuals are lower at flows higher than 35 $m^3 s^{-1}$ compared to at flows less than 35 $m^3 s^{-1}$ in Fig. 6a, it can be seen from Fig. 6b that the HBV-96 model is less accurate in simulating high flows compared to low flows.

Figure 7a presents probability plots of model residuals comparing the two selected distributions (normal distribution and t location-scale distribution). The estimated parameters for the best fit to data are $\mu = 0.0363$ and $\sigma = 0.7619\,m^3 s^{-1}$ for normal distribution – same with the empirical parameters. On the other hand, the best fit parameters for t location-scale distribution are different: $\mu = 0.0607\,m^3 s^{-1}$, $\sigma = 0.2351\,m^3 s^{-1}$ and $\nu = 1.5833$. From Fig. 7a, one can conclude that the model residuals' distribution is far from being close to normal even though the parameters of the fitted normal distribution are the same with those obtained from the empirical distribution. It is obvious that t location-scale distribution provides better fit as it is able to enclose the data at the tails much better compared to fitted normal distribution. Yet, outliers are still not represented fully.

Normality of model residuals' distribution is further investigated for different hydrometeorological conditions as identified by clustering in the space of the predictor variables. Analysis of probability plot for each cluster formed indicates that there is no significant departure from normality (with regard to the fitted normal distribution) unlike in the overall model residuals. The most striking result among all clusters is achieved in the one representing very high flow and high rainfall (0.95 % of total data) (Fig. 7b). It should be noted that it is mostly these extreme events making overall residuals distribution non-Gaussian. Classifying data so that different hydrometeorological conditions, most importantly extreme events, are separated helps to achieve homogeneity, and thus normality in model residuals' distribution. Therefore clustering can be suggested as an alternative to transformation of model residuals before applying any statistical methods on them.

4.1.2 Upper Severn catchments: Yeaton, Llanyblodwel, and Llanerfyl

Understanding the quality of (water level) forecasts is important in order to efficiently discuss uncertainty analysis results provided by any method. In Upper Severn catchments, this is done based on standard deviation of model error. The results are comparatively presented for different lead times in Fig. 8 where the effect of lead time on forecast quality can be clearly seen. As lead time increases, the standard deviation of error increases as well. Also, it should be noticed that there is a direct increasing effect of shorter basin lag time on standard deviation. For example, catchment with shortest basin lag time, that is Llanerfyl, has always larger standard deviation for all lead times. On the contrary, the smallest standard deviation always occurs in the catchment having the longest basin lag time, which is Yeaton. This is mainly due to the fact that the basin lag time represents memory of a catchment. Hence, flood forecasting capability of a hydrological model is affected negatively when the basin lag time is short.

HESSD
11, 10179–10233, 2014

The observed water levels are plotted against forecasted water levels in Llanyblodwel catchment for lead time = 6 h in Fig. 9a. Figure 9b shows model error plotted against observed water level on the logarithmic scale. Although it is not very clear from Fig. 9a, it is evident from Fig. 9b that the model error increases with higher water levels, as expected.

Normality of model residuals for Llanyblodwel catchment for all lead times was investigated (see Fig. 10a). Visual inspection of probability plots, superimposed on which the line joining the 25th and 75th percentiles of the fitted normal distributions, reveals that errors are not normally distributed, i.e. the data does not fall on the straight line as it is especially the case for the tails. It should be realized that the departure from normality increases with longer lead times.

Furthermore, a normality check for model residuals' distribution is made individually for the data clusters corresponding to particular hydrometerological conditions. The variables used for clustering are groundwater level ($\boldsymbol{GW}$), soil moisture deficit ($\boldsymbol{SMD}$), and observed water level ($\boldsymbol{H}_{\text{obs},t-1}$). It is seen that the level of achieving normality in model residuals' distribution for each cluster is substantially poorer if compared to the Brue catchment. This can be explained by the fact that the error time series data being analyzed has a time step of 12 h which is long enough to hinder effects of varying water levels on error. Another reason can be related to the nature of model residuals, e.g. forecasted precipitation is used to predict water levels. This brings a great amount of uncertainty and a higher difference between the actual and the predicted water levels (i.e. higher model residuals). It is also worth mentioning that the distribution closest to normal is found in the data cluster representing high groundwater levels, high water levels, and low soil moisture deficit (4.6 % of the total data set) (Fig. 10b).

HESSD
11, 10179–10233, 2014

4.2 Uncertainty prediction by the two methods

4.2.1 Brue catchment

Validation measures PICP, MPI, and ARIL are provided in Table 2. In terms of PICP, even though QR provides PICP values slightly closer to 90 % and 50 % during training, UNEEC was found to be more reliable in validation especially for the 90 % confidence level. While the narrowest prediction interval on average is given by UNEEC during training for both 90 % and 50 % confidence level, comparable MPI values are obtained during validation. QR stands outs with its smaller ARIL values particularly for the 90 % confidence level. However, if one considers PICP and MPI, as well as ARIL, on aggregate UNEEC yields better results over QR.

Looking at Fig. 11a, visual analysis of 90 % prediction intervals for the highest flow period in validation reveals that neither UNEEC nor QR is perfectly able to enclose peak observations of high flows. Overall, uncertainty analysis results from UNEEC and QR are comparable. Yet, in comparison to UNEEC, QR produces unnecessarily wider uncertainty bounds for medium peaks in validation (see Fig. 11b). The reason for this can be related to the fact that these medium peaks last longer. UNEEC is able to memorize catchment behaviour far better as it considers encapsulated information of catchment characteristics in its multiple predictors.

PICP, MPI, ARIL and NUE values for each cluster are computed for QR and UNEEC. The results are listed in Table 3. From this table, it is possible to verify the contradictory relationship between PICP and MPI when the two methods are compared: PICP is closer to 90 % more when MPI is higher. Unlike for the whole data set (that is highly heterogeneous due to extremes in rainfall–runoff process), such relationship is observed when the homogeneous data sets (e.g. clusters) are analyzed. It is likely that considering all the data having varying uncertainty width over the available period compensates peculiarities of each cluster regarding their own PICP and MPI. (Note that ARIL also has a similar situation.) Based on this explanation, comparison of the methods QR and UNEEC for the cluster of high flow (and high rainfall) cannot be made properly when

HESSD
11, 10179–10233, 2014

one considers PICP, MPI, and ARIL only. Therefore, NUE values ($w = 1$) corresponding to each cluster are also computed. Based on NUE values, it can be concluded that in general UNEEC shows a better performance since it yields a higher NUE value (except for the first cluster).

4.2.2 Upper Severn catchments: Yeaton, Llanyblodwel, and Llanerfyl

MPI and ARIL values for the 90 % confidence level on validation data set are presented in Fig. 12. The effect of lead time on both measures of uncertainty is such that when the lead time increases, quality of the forecast decreases, hence the values of both measures increases. In view of the model quality for predicted water levels, the relatively low MPI values in Yeaton catchment are not surprising for both methods. Having the longest basin lag time, accuracy of forecast in Yeaton catchment is much higher. On the other hand, the catchment with the shortest basin lag time, that is Llanerfyl, always has the largest MPI. The following points need mentioning if the two methods are compared:

- In terms of MPI: (i) QR gives slightly narrower bands for Yeaton catchment especially at longer lead times; (ii) the methods perform equally well for Llanyblodwel catchment at all lead times; (iii) UNEEC provides relatively lower MPI values than QR for Llanerfyl.

- Based on ARIL: (i) while QR outperforms UNEEC in Yeaton especially for longer lead times, the methods provide nearly the same values for Llanyblodwel and Llanerfyl. It should be noticed that, for Llanerfyl, ARIL values from UNEEC method are always the smallest.

MPI values plotted against PICP values for the validation period are shown in Fig. 13. Overall, when one considers both MPI and PICP:

- Yeaton: QR does slightly better than UNEEC. Low PICP values obtained by UNEEC at shorter lead times are due to higher predictive accuracy of the forecasting

HESSD
11, 10179–10233, 2014

model (i.e., model residuals less than 1 mm). It should be noted that in this study QR method, unlike UNEEC, predicts the quantiles of the uncertain water level rather than of the residual error. Such an approach eliminates the possibility of having extremely low PICP values resulting from the cases where the model is able to predict the variable of interest quite well.

– Llanyblodwel: Both methods are equally capable of providing reasonably well uncertainty estimates (as measured by both MPI and PICP).

– Llanerfyl: UNEEC method is outperforming QR method in terms of both MPI and PICP.

For further comparison of estimated prediction limits through uncertainty plots, three cases are selected based on the relationship between basin lag time and lead time. These cases are (1) Yeaton, lead time = 3 h (lead time < basin lag time), (2) Llanyblodwel, lead time = 6 h (lead time ≈ basin lag time), and (3) Llanerfyl, lead time = 12 h (lead time > basin lag time). The fundamental idea here is to understand how well the residual uncertainty is assessed with regard to forecast lead time and its relation to basin lag time. The catchment with the longest basin lag time (Yeaton) is considered for Case 1, where the effect of a very short lead time is to be investigated. Here on this decision, there is the deliberate intention to combine the condition of having more accurate model outputs (i.e. extremely small residuals) as well. Case 3, on the other hand, is important to understand lead time-basin lag time relationship for the worst situation: relatively poor quality of forecasting model and the longest lead time. This is the most critical case in that the performance of predictive uncertainty method's performance has a bigger role in operational decision making process. Apart from these two extreme cases, Case 2 represents a balanced situation where the lead time of interest and basin lag time are approximately equal. Llanyblodwel catchment is chosen for this case as its model has a moderate predictive accuracy. Figure 14 compares the computed prediction limits by QR and UNEEC for these cases during the latest 11 months period of validation (April 2012–February 2013). It was during late 2012 that

HESSD
11, 10179–10233, 2014

Upper Severn catchment suffered from serious flooding and this period corresponds to the right half of the plots. The most salient observations from Fig. 14 are as follows:

- In Llanerfyl, one can notice a strange behaviour of the model causing sharp changes in forecasted water levels (unstable model outputs), and thus in prediction limits. Considering that Llanerfyl catchment has a basin lag time of $\sim$ 3–5 h, hydrological conditions in the catchment, e.g. water levels, can change significantly in 12 h (Δt, time step of the data set). Therefore, it is not surprising that the sharpest changes occur in this catchment's hydrograph as compared to Yeaton and Llanyblodwel. One can observe even more significant changes in the second half period of the hydrograph. It is necessary to mention that these oscillating changes appear as a consequence of the forecasting model's extremely poor performance.

- For medium water levels in Yeaton and Llanybldowel, UNEEC gives wider prediction intervals as compared to QR, particularly on falling limb part of the hydrographs. A possible explanation for this can be encapsulation of groundwater level information in UNEEC. Groundwater levels remains at higher levels for longer periods than water levels in the river (i.e. due to slow and long response time of groundwater levels to changing hydrometeorological conditions) and thus UNEEC has the potential to provide uncertainty band of larger widths.

- For peak water levels in Yeaton and Llanyblodwel catchments, it is mostly QR that produces higher upper prediction limit than UNEEC. Yet, this doesn't contribute to overall performance of the method significantly. On the contrary, it is seen in some cases that such high upper prediction limits makes the uncertainty band unnecessarily too wide.

- Continuous peaks prevail in Llanerfyl catchment (as its basin lag time is way shorter than the forecast lead time of interest). Such continuous peaks occur during certain periods in Llanyblodwel catchment too. In most of these cases,

HESSD
11, 10179–10233, 2014

UNEEC gives narrower uncertainty band, and wider prediction interval computed by QR is redundant. That is to say, it doesn't contribute QR method's performance (as measured by PICP) at all in terms of its ability to enclose more observations within the band. For peak water levels, however, QR is slightly more informative than UNEEC.

– Noticeably, upper prediction limits obtained by QR in Llanerfyl catchment for the long-lasting falling limb part of the hydrograph (indicated by arrows in Fig. 14c) are too high, e.g. even greater than those provided by UNEEC. QR is a method building simple linear regression models considering only observed water levels on forecasted water levels. Having rather simple mathematical formulation, it might be that sensitivity of the computed upper prediction limit to magnitude of water level increases, and shows an amplifying effect on uncertainty band width.

Table 4 shows the values of validation measures (MPI, PICP, and ARIL) for each cluster for Llanyblodwel catchment (lead time = 6 h). In UNEEC, the highest MPI value was obtained for cluster 2 (highest groundwater levels) with a relatively bad PICP value compared to other clusters. The low PICP in cluster 2 can be explained by limited number of data (only 4.6 %) available for highest groundwater levels occurring rarely. Similar to UNEEC, the highest MPI was also obtained for this cluster with QR method. Providing a wider uncertainty band than UNEEC on average, QR is not very much capable of estimating reasonable prediction limits for very high groundwater levels. This is also supported by its greater (12 %) ARIL value compared to UNEEC.

PICP and MPI values for the cluster 4 should be mentioned as well. This cluster represents very low water levels, very low groundwater levels, and very high soil moisture deficit, and constitutes 16.6 % of the whole data. As distinct from cluster 2, bad (but slightly better) PICP value (obtained by UNEEC) in cluster 4 can be attributed to its lower MPI. In comparison to UNEEC, QR provides PICP values which are close to target value (i.e. 90 %) despite its lower MPI. Thus, one can say that UNEEC certainly fails in providing reliable uncertainty estimates for the extreme condition associated to

HESSD
11, 10179–10233, 2014

very low water and groundwater levels. This can be due to the effect of using state variables as predictors. All in all, state variables are model outputs and they cannot reflect real catchment conditions truly especially when the (hydrological) model is not very accurate. That is particularly true for the extreme events considering that models mostly fail in simulating such events.

5 Conclusions and recommendations

This study should be seen as accompanying the study by López López et al. (2014) (and earlier work on UNEEC and QR) and presents a comparative evaluation of uncertainty analysis and prediction results from QR and UNEEC methods on the four catchments that vary in its hydrological characteristics and models: Brue catchment (simulation mode) and Upper Severn catchments – Yeaton, Llanyblodwel, and Llanerfyl (forecasting mode). The latter set of case studies is important from a practical perspective in that the effect of lead time on uncertainty analysis results and its relation with basin lag time is demonstrated. For both QR and UNEEC different model configurations than their previous applications are considered. The following conclusions can be drawn from the results of this study:

- In terms of easiness of setup (data preparation and calibration), preference should be given to QR simply because it is a simpler linear method with one input variable (in this study), whereas UNEEC has more steps and requires more data analysis. However, this has to be done only once (i.e. during calibration), and in operation both methods can be easily used and have equally low running times (a fraction of a second) since they are based on algebraic calculations.

- In almost all case studies both methods adequately represent residual uncertainty and provide similar results consistent with understanding of the hydrological picture of the catchment and the accuracy of the (hydrological) models used. We can recommend both methods for the use in flood forecasting.

HESSD

11, 10179–10233, 2014

- It is impossible to have a strict validation of predictive uncertainty methods on real life data since there is no basis for comparison, i.e. the values of residual error quantiles are unknown. In this study, several measures (PICP, MPI, ARIL, and NUE) that jointly provide a certain indication of the methods' quality have been used. The methods have been also compared based on the empirical judgement about the residual uncertainty under different hydrometeorological conditions and different states of a catchment in terms of the hydrological response.

- In one case study, Llanerfyl, we found that UNEEC was giving more adequate estimates than QR. This catchment has a shorter basin lag time and the model outputs for this catchment were characterized by a relatively high error, so our conclusion was that probably in such a rapid response catchment the UNEEC's more sophisticated non-linear models were able to capture relationships between quantiles, and hydrometeorological and state variables better than the QR's linear model. Introducing more predictors in the QR methodology (thus "pushing" towards UNEEC) could possibly increase the performance of QR for Llanerfyl.

- A useful finding is that inclusion of a variable representing groundwater level (GW) as a predictor in UNEEC improves its performance for the Upper Severn catchments. This can be explained by the fact that this variable has a high level of information content about the state of a catchment. However, it should be noted that in other catchments using such information can be misleading due to slow (and long) response time of groundwater levels to changing hydrometeorological conditions. Yet, overall, it can be advised to make use of variables which can be representative of hydrological response behaviour of a catchment for improving the predictive capacity of the data-driven methods.

We recommend comparing the two presented methods (QR and UNEEC) with more predictive uncertainty methods which use different methodologies, such as HUP (Krzysztofowicz, 1999) or DUMBRAE (Pianosi and Raso, 2012). It is also necessary to test capabilities of different predictive uncertainty methods on catchments from regions of distinct hydrologic behaviour, with diverse climatic conditions, and having various hydrological features. In this study, we found that the basin lag time is a notable characteristic of a catchment having great influence on uncertainty analysis results (as measured by MPI and PICP). When the lag time is longer, the catchment memorizes more information regarding its hydrological response characteristics.

On the other hand, exploring the performance of different methods on *similar catchments* (Sawicz et al., 2011; Toth, 2013; Patil and Stieglitz, 2011; Sivakumar et al., 2014) and finding bases for generalized guidelines on the selection of most appropriate predictive uncertainty method in operational flood forecasting practices is also important and could be considered in the further studies as well.

When different predictive uncertainty methods are evaluated based on their comparative performance, it is more important to have validation measures incorporating certain aspects of rainfall–runoff process, i.e. varying flow conditions. For example, the accuracy of the hydrological model, thus the amount of residual uncertainty, decreases during high flow values. This necessitates exploring validation measures linking prediction interval to the (hydrological) model quality (see, e.g. Dogulu et al., 2014: weighted mean prediction interval).

We also see other possibilities for further improvements in both methods. For example, the different configurations of QR and alternative clustering techniques for UNEEC can be explored further.

Acknowledgements. The results presented in this paper were obtained during thesis research of European Erasmus Mundus Flood *Risk* Master students Nilay Dogulu and Patricia López López. The Environment Agency is gratefully acknowledged for providing the data for Upper Severn catchments in the UK.

HESSD

11, 10179–10233, 2014

References

Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfallrunoff model, J. Hydrol., 280, 145–161, doi:10.1016/S0022-1694(03)00229-4, 2003.

Bailey, R. and Dobson, C.: Forecasting for floods in the Severn catchment, J. Inst. Water Eng. Sci., 35, 168–178, 1981.

Barnwal, P. and Kotani, K.: Climatic impacts across agricultural crop yield distributions: an application of quantile regression on rice crops in Andhra Pradesh, India, Ecol. Econ., 87, 95–109, 2013.

Battiti, R.: Using mutual information for selecting features in supervised neural net learning, IEEE T. Neural Networ., 5, 537–550, 1994.

Baur, D., Saisana, M., and Schulze, N.: Modelling the effects of meteorological variables on ozone concentration – a quantile regression approach, Atmos. Environ., 38, 4689–4699, 2004.

Bell, V. A. and Moore, R. J.: The sensitivity of catchment runoff models to rainfall data at different spatial scales, Hydrol. Earth Syst. Sci., 4, 653–667, doi:10.5194/hess-4-653-2000, 2000.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., and Murtagh, R. F.: Validity-guided (re) clustering with applications to image segmentation, IEEE T. Fuzzy Syst., 4, 112–123, 1996.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments Rep. RHO7, 118 Swed. Meteorol. and Hydrol. Inst., Norrköping, Sweden, 1976.

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, doi:10.1002/hyp.3360060305, 1992.

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

Bondell, H. D., Reich, B. J., and Wang, H.: Noncrossing quantile regression curve estimation, Biometrika, 97, 825–838, doi:10.1093/biomet/asq048, 2010.

Bremnes, J. B.: Probabilistic wind power forecasts using local quantile regression, Wind Energy, 7, 47–54, doi:10.1002/we.107, 2004.

Brown, J. D. and Heuvelink, G. B. M.: Assessing uncertainty propagation through physically based models of soil water flow and solute transport, in: Encyclopedia of Hydrological Sciences, 6: 79, edited by: Anderson, M. G., John Wiley, New York, 2005.
Cannon, A. J.: Quantile regression neural networks: implementation in R and application to precipitation downscaling, Comput. Geosci., 37, 1277–1284, 2011.
Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrol. Earth Syst. Sci., 15, 3253–3274, doi:10.5194/hess-15-3253-2011, 2011.
Coxon, G., Freer, J., Westerberg, I., Woods, R., Smith, P., and Wagener, T.: A generalised framework for large-scale evaluation of discharge uncertainties across England and Wales, Geophysical Research Abstracts, 16, EGU2014-10157, EGU General Assembly 2014, Vienna, Austria, 2014.
Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, doi:10.5194/hess-13-913-2009, 2009.
Dogulu, N., Solomatine, D. P., and Shrestha, D. L.: Applying clustering approach in predictive uncertainty estimation: a case study with the UNEEC method, Geophysical Research Abstracts, 16, EGU2014-5992, EGU General Assembly 2014, Vienna, Austria, 2014.
EA: Environment Agency: River levels: Midlands, available at: http://www.environment-agency.gov.uk/homeandleisure/floods/riverlevels/, last access: 1 October 2013.
Friederichs, P. and Hense, A.: Statistical downscaling of extreme precipitation events using censored quantile regression, Mon. Weather Rev., 135, 2365–2378, doi:10.1175/MWR3403.1, 2007.
Gupta, H. V., Beven, K. J., and Wagener, T.: Model calibration and uncertainty estimation, in: Encyclopedia of Hydrological Sciences, 11:131, edited by: Anderson, M. G., John Wiley, New York, 2005.
Halkidi, M., Batistakis, Y., and Vazirgiannis, M.: On clustering validation techniques, J. Intell. Inf. Syst., 17, 107–145, doi:10.1023/A:1012801612483, 2001.
Hill, T. and Neal, C.: Spatial and temporal variation in pH, alkalinity and conductivity in surface runoff and groundwater for the Upper River Severn catchment, Hydrol. Earth Syst. Sci., 1, 697–715, doi:10.5194/hess-1-697-1997, 1997.
Jin, X., Xu, C. Y., Zhang, Q., and Singh, V. P.: Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model, J. Hydrol., 383, 147–155, doi:10.1016/j.jhydrol.2009.12.028, 2010.

Kobold, M. and Sušelj, K.: Precipitation forecasts and their uncertainty as input into hydrological models, Hydrol. Earth Syst. Sci., 9, 322–332, doi:10.5194/hess-9-322-2005, 2005.
Koenker, R.: Quantile Regression, Cambridge University Press, New York, 2005.
Koenker, R. and Bassett Jr., G.: Regression quantiles, Econometrica, 1, 33–50, 1978.
Koenker, R. and Hallock, K.: Quantile regression, J. Econ. Perspect., 15, 143–156, 2001.
Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resour. Res., 35, 2739–2750, doi:10.1029/1999WR900099, 1999.
Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, J. Hydrol., 249, 2–9, 2001.
Kudryavtsev, A. A.: Using quantile regression for rate-making, Insur. Math. Econ., 45, 296–304, doi:10.1016/j.insmatheco.2009.07.010, 2009.
Lee, H., Seo, D.-J., and Koren, V.: Assimilation of streamflow and in-situ soil moisture data into operational distributed hydrologic models: effects of uncertainties in the data and initial model soil moisture states, Adv. Water Resour., 34, 1597–1615, doi:10.1016/j.advwatres.2011.08.012, 2011.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, doi:10.1016/S0022-1694(97)00041-3, 1997.
Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: toward an integrated data assimilation framework, Water Resour. Res., 43, W07401, doi:10.1029/2006WR005756, 2007.
López López, P., Verkade, J. S., Weerts, A. H., and Solomatine, D. P.: Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: a comparison, Hydrol. Earth Syst. Sci., 18, 3411–3428, doi:10.5194/hess-18-3411-2014, 2014.
Marsh, T. and Hannaford, J.: UK Hydrometric Register, Hydrological Data UK Series, Centre for Ecology and Hydrology, Wallingford, UK, 1–210, 2008.
Møller, J. K., Nielsen, H. A., and Madsen, H.: Time-adaptive quantile regression, Comput. Stat. Data An., 52, 1292–1303, 2008.
Montanari, A.: What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Proc., 21, 841–845, 2007.
Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall–runoff simulations, Water Resour. Res., 40, W01106, doi:10.1029/2003WR002540, 2004.

HESSD
11, 10179–10233, 2014

Moore, R. J., Jones, D. A., Cox, D. R., and Isham, V. S.: Design of the HYREX raingauge network, Hydrol. Earth Syst. Sci., 4, 521–530, doi:10.5194/hess-4-521-2000, 2000.
Mukolwe, M. M., Di Baldassarre, G., Werner, M., and Solomatine, D. P.: Flood modelling: parameterisation and inflow uncertainty, P. I. Civil Eng.-Wat. M., 167, 51–60, 2014.
Munir, S., Chen, H., and Ropkins, K.: Modelling the impact of road traffic on ground level ozone concentration using a quantile regression approach, Atmos. Environ., 60, 283–291, 2012.
Nasseri, M. and Zahraie, B.: Application of simple clustering on space-time mapping of mean monthly rainfall pattern, Int. J. Climatol., 31, 732–741, doi:10.1002/joc.2109, 2011.
Nasseri, M., Zahraie, B., Ansari, A., and Solomatine, D. P.: Uncertainty assessment of monthly water balance models based on Incremental Modified Fuzzy Extension Principle method, J. Hydroinform., 15, 1340–1360, doi:10.2166/hydro.2013.159, 2013.
Nielsen, H. A., Madsen, H., and Nielsen, T. S.: Using Quantile Regression to extend an existing wind power forecasting system with probabilistic forecasts, Wind Energy, 9, 95–108, 2006.
Pappenberger, F. and Beven, K. J.: Ignorance is bliss: or seven reasons not to use uncertainty analysis, Water Resour. Res., 42, W05302, doi:10.1029/2005WR004820, 2006.
Patil, S. and Stieglitz, M.: Hydrologic similarity among catchments under variable flow conditions, Hydrol. Earth Syst. Sci., 15, 989–997, doi:10.5194/hess-15-989-2011, 2011.
Pianosi, F. and Raso, L.: Dynamic modeling of predictive uncertainty by regression on absolute errors, Water Resour. Res., 48, W03516, doi:10.1029/2011WR010603, 2012.
Pianosi, F., Shrestha, D. L., and Solomatine, D. P.: ANN-based representation of parametric and residual uncertainty of models, IEEE IJCNN, 1–6, doi:10.1109/IJCNN.2010.5596852, 2010.
Quinlan, J. R.: Learning with continuous classes, in: Proceedings of the 5th Australian joint Conference on Artificial Intelligence, Hobart, Tasmania, 16–18 November 1992, 343–348, 1992.
Reggiani, P. and Weerts, A.: A Bayesian approach to decision-making under uncertainty: an application to real-time forecasting in the river Rhine, J. Hydrol., 356, 56–59, doi:10.1016/j.jhydrol.2008.03.027, 2008.
Reggiani, P., Renner, M., Weerts, A., and Van Gelder, P.: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, Water Resour. Res., 45, W02428, doi:10.1029/2007WR006758, 2009.
Roscoe, K. L., Weerts, A. H., and Schroevers, M.: Estimation of the uncertainty in water level forecasts at ungauged river locations using Quantile Regression, Int. J. River Basin Manage., 10, 383–394, 2012.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.
Shrestha, D. L. and Solomatine, D. P.: Machine learning approaches for estimation of prediction interval for the model output, Neural Networks, 19, 225–235, doi:10.1016/j.neunet.2006.01.012, 2006.
Shrestha, D. L. and Solomatine, D. P.: Data-driven approaches for estimating uncertainty in rainfall–runoff modelling, Int. J. River Basin Manage., 6, 109–122, 2008.
Shrestha, D. L., Rodriguez, J., Price, R. K., and Solomatine, D. P.: Assessing model prediction limits using fuzzy clustering and machine learning, in: Proc. 7th Int. Conf. On Hydroinformatics, 4–8 September, Acropolis, Nice, France, 2006.
Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., and Hapuarachchi, H. A. P.: Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose, Hydrol. Earth Syst. Sci., 17, 1913–1931, doi:10.5194/hess-17-1913-2013, 2013.
Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J.: Considering rating curve uncertainty in water level predictions, Hydrol. Earth Syst. Sci., 17, 4415–4427, doi:10.5194/hess-17-4415-2013, 2013.
Singh, S. K., McMillan, H., and Bárdossy, A.: Use of the data depth function to differentiate between case of interpolation and extrapolation in hydrological model prediction, J. Hydrol., 477, 213–228, doi:10.1016/j.jhydrol.2012.11.034, 2013.
Sivakumar, B., Singh, V. P., Berndtsson, R., and Khan, S. K.: Catchment Classification Framework in Hydrology: Challenges and Directions, J. Hydrol. Eng., doi:10.1061/(ASCE)HE.1943-5584.0000837, in press, 2014.
Solomatine, D. P.: Modelling Theory and Uncertainty (An Introduction) – Part 2: Analysis of Model Uncertainty, Lecture Notes, UNESCO-IHE Institute for Water Education, Delft, the Netherlands, 2013.
Solomatine, D. P. and Shrestha, D. L.: A novel method to estimate model uncertainty using machine learning techniques, Water Resour. Res., 45, W00B11, doi:10.1029/2008WR006839, 2009.
Solomatine, D. P. and Wagener, T.: Hydrological Modeling, in: Treatise on Water Science, vol. 2, edited by: Wilderer, P., Academic Press, Oxford, 435–457, 2011.

Solomatine, D. P., Dibike, Y. B., and Kukuric, N.: Automatic calibration of groundwater models using global optimization techniques, Hydrol. Sci. J., 44, 879–894, 1999.
Tayfur, G., Zucco, G., Brocca, L., and Moramarco, T.: Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale, J. Hydrol., 510, 363–371, doi:10.1016/j.jhydrol.2013.12.045, 2014.
Taylor, J. W.: Forecasting daily supermarket sales using exponentially weighted quantile regression, Eur. J. Oper. Res., 178, 154–167, 2007.
Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Int. J. River Basin Manage., 6, 123–137, 2008.
Toth, E.: Catchment classification based on characterisation of streamflow and precipitation time series, Hydrol. Earth Syst. Sci., 17, 1149–1159, doi:10.5194/hess-17-1149-2013, 2013.
van Andel, S. J., Weerts, A., Schaake, J., and Bogner, K.: Post-processing hydrological ensemble predictions intercomparison experiment, Hydrol. Process., 27, 158–161, doi:10.1002/hyp.9595, 2013.
Verkade, J. S. and Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning, Hydrol. Earth Syst. Sci., 15, 3751–3765, doi:10.5194/hess-15-3751-2011, 2011.
Wagener, T. and Gupta, H. V.: Model identification for hydrological forecasting under uncertainty, Stoch. Env. Res. Risk A., 19, 378–387, doi:10.1007/s00477-005-0006-5, 2005.
Wallingford: Wallingford Water, a Flood Forecasting and Warning System for the River Soar, Wallingford Water, Wallingford, UK, 1994.
Wallingford: HR Wallingford, ISIS software, available at: http://www.isisuser.com/isis/, HR Wallingford, Hydraluic Unit, Wallingford, UK, 1997, last access: 1 October 2013.
Weerts, A. H., Winsemius, H. C., and Verkade, J. S.: 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, doi:10.5194/hess-15-255-2011, 2011.
Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., and Heynert, K.: The Delft-FEWS flow forecasting system, Environ. Modell. Softw., 40, 65–77, 30 doi:10.1016/j.envsoft.2012.07.010, 2013.
Xie, X. L. and Beni, G.: A validity measure for fuzzy clustering, IEEE T. Pattern. Anal., 13, 841–847, 1991.

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Table 1. Summary of the main basin characteristics for the catchments selected.

Catchment name	Drainage area (km^2)	Elevation (m)	Mean flow ($m^3 s^{-1}$)	Mean annual rainfall (mm)	Basin lag time (h)
Brue	135	≈ 20	1.92[a]	867[a]	8–9
Yeaton	180.8	61.18	1.60[b]	767[b]	15–20
Llanyblodwel	229	77.28	6.58[b]	1267[b]	7–10
Llanerfyl	≈ 100	151	> 10[c]	> 1300[c]	3–5

[a] Basin average for the period 1961–1990.
[b] Computed for the periods 1963–2005 and 1973–2005 for Yeaton and Llanyblodwel, respectively and taken from UK Hydrometric Register (Marsh and Hannaford, 2008).
[c] Rough estimates based on the data available for 2006–2013.

HESSD
11, 10179–10233, 2014

Table 2. Uncertainty analysis results for 90 % and 50 % confidence levels (Brue catchment).

Confidence level		PICP (%)		MPI ($m^3 s^{-1}$)		ARIL (–)	
		UNEEC	QR	UNEEC	QR	UNEEC	QR
TR	90 %	91.19	90.00	1.58	1.69	1.86	1.47
	50 %	51.28	50.01	0.54	0.58	0.55	0.46
VD	90 %	88.29	82.33	1.37	1.39	2.35	1.83
	50 %	30.29	32.75	0.45	0.47	0.67	0.57

TR: Training, VD: Validation.

HESSD
11, 10179–10233, 2014

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Table 3. PICP, MPI, ARIL and NUE values for each cluster (training, 90 % confidence level, Brue): UNEEC vs. QR.

Cluster No	Number of data	UNEEC PICP (%)	UNEEC MPI ($m^3 s^{-1}$)	UNEEC ARIL (–)	UNEEC NUE (–)	QR PICP (%)	QR MPI ($m^3 s^{-1}$)	QR ARIL (–)	QR NUE (–)
1[a]	5447 (62.3%)	92.12	1.14	2.67	34.5	88.16	0.88	1.96	45.0
2	787 (9.0 %)	82.08	2.98	0.50	164.2	84.50	3.51	0.57	148.2
3	2167 (24.7%)	94.46	1.44	0.53	178.2	96.72	1.94	0.71	136.2
4[b]	83 (0.95 %)	**74.70**	**7.55**	**0.33**	**226.4**	**90.36**	**12.00**	**0.49**	**184.4**
5	266 (3.05 %)	77.44	5.96	0.48	161.3	89.47	7.58	0.58	154.3

[a] Low flows, low rainfall.
[b] High flows, high rainfall.

Table 4. PICP, MPI, and ARIL values for each cluster (training, 90 % confidence level, Llanyblodwel, lead time = 6 h): UNEEC vs. QR.

Cluster No	Number of data	UNEEC PICP (%)	UNEEC MPI ($m^3 s^{-1}$)	UNEEC ARIL (–)	QR PICP (%)	QR MPI ($m^3 s^{-1}$)	QR ARIL (–)
1	413 (19.1%)	88.62	0.1492	0.271	93.95	0.1506	0.250
2[a]	**100 (4.6%)**	**85.00**	**0.2964**	**0.288**	**95.00**	**0.3538**	**0.326**
3	336 (15.5 %)	90.18	0.1798	0.249	94.94	0.2283	0.287
4[b]	**359 (16.6%)**	**93.04**	**0.0518**	**0.182**	**89.14**	**0.0305**	**0.100**
5	535 (24.8%)	89.53	0.1128	0.308	85.79	0.0742	0.179
6	416 (19.2%)	90.38	0.0920	0.212	92.31	0.1021	0.208

[a] High groundwater levels.
[b] Low groundwater levels.

HESSD
11, 10179–10233, 2014

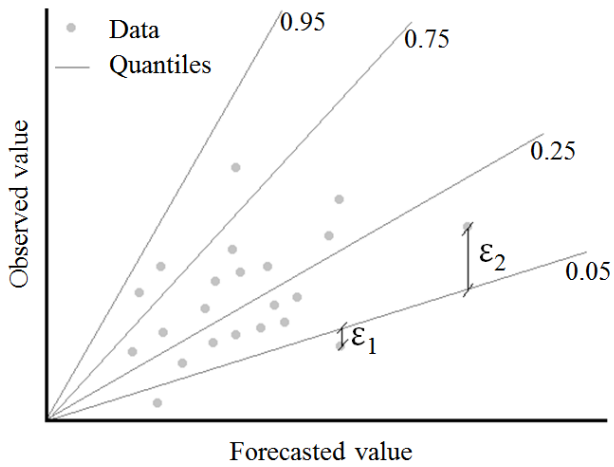

Figure 1. Quantile regression example scheme considering different quantiles.

HESSD
11, 10179–10233, 2014

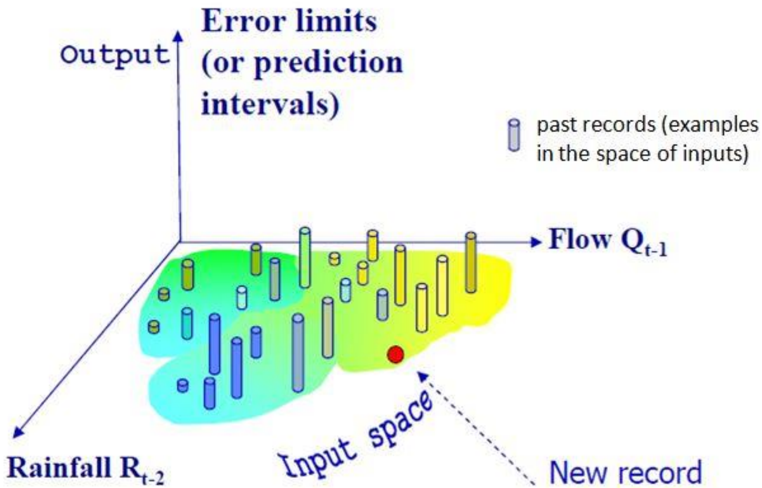

Figure 2. An example to fuzzy clustering of input data (the predictors are past rainfall at lag $t-2$ and past flow at lag $t-1$) during training of the uncertainty model, U (adapted from Solomatine, 2013).

HESSD
11, 10179–10233, 2014

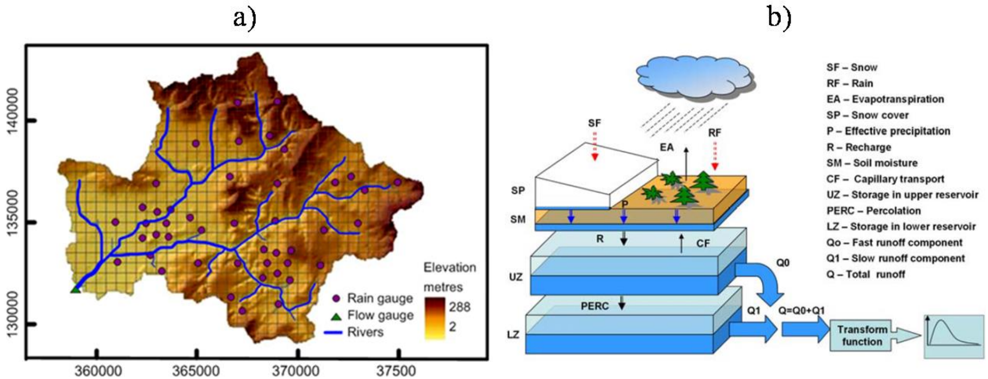

Figure 3. (a) The Brue catchment showing dense rain gauges network and its river gauging station, Lovington, where the discharge is measured, and **(b)** schematic representation of HBV-96 model (Lindström et al., 1997) with routine for snow (upper), soil (middle), and response (bottom) (Shrestha and Solomatine, 2008).

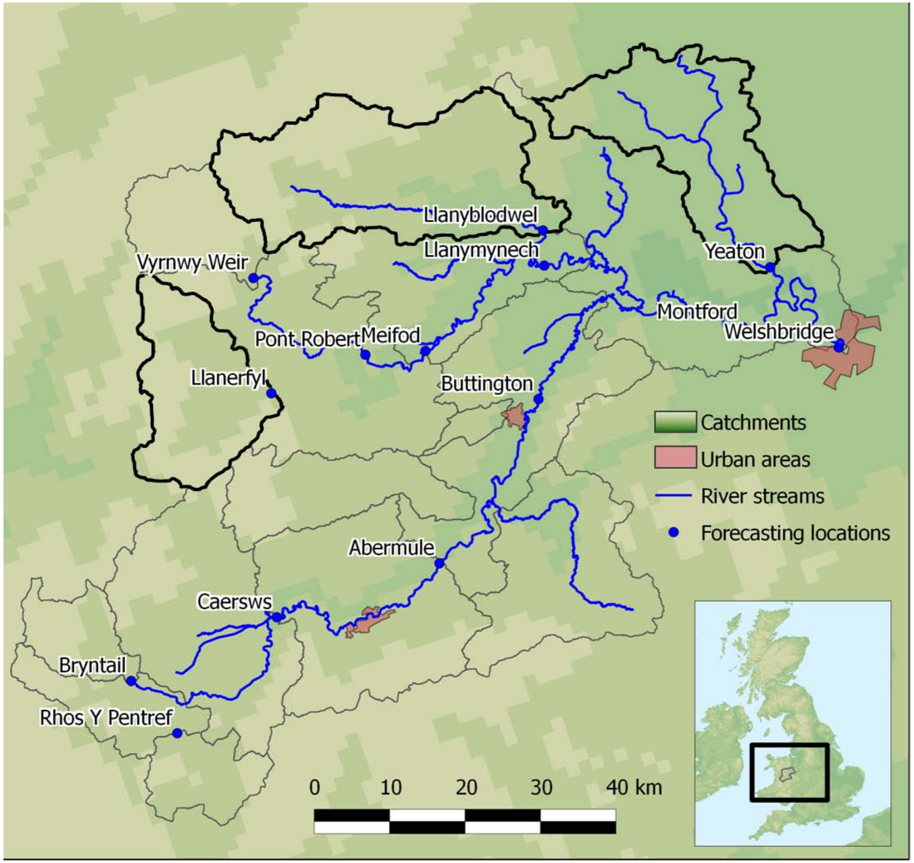

Figure 4. The Upper Severn catchments: Yeaton, Llanyblodwel and Llanerfyl (adapted from López López et al., 2014).

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀I ▶I
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

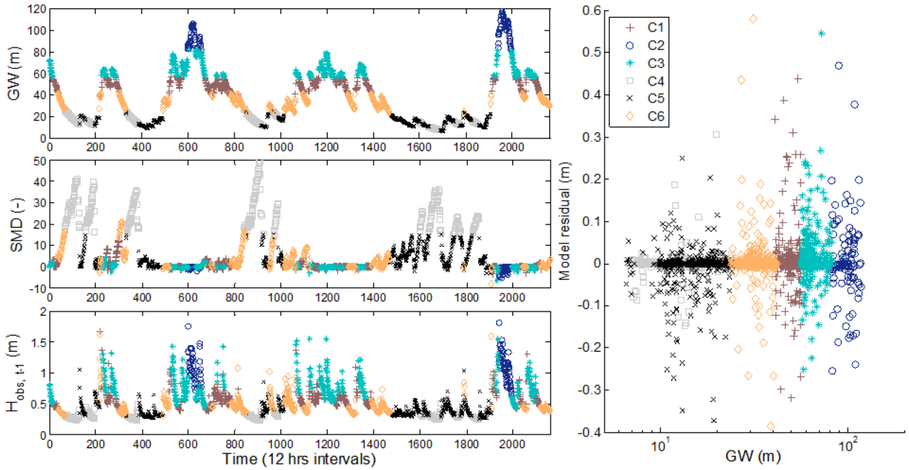

Figure 5. Fuzzy clustering of: ***GW*** (left, top) and its relation with the model residuals (right), ***SMD*** (left, middle) and $\boldsymbol{H}_{\mathrm{obs},t-12}$, (left, bottom) for calibration period (7 March 2007 08:00–7 March 2010 08:00) – Llanyblodwel, lead time = 6 h.

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

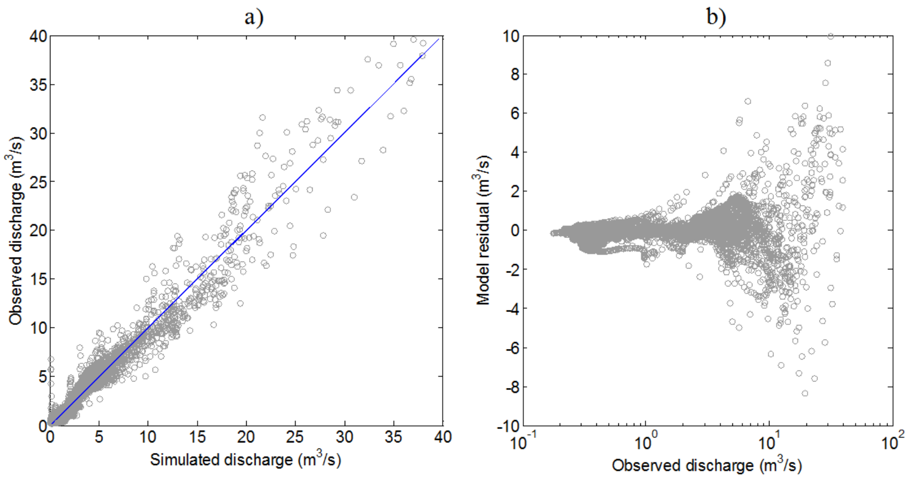

Figure 6. Observed discharge, simulated discharge, and model residuals during calibration (Brue catchment).

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂◂ | ▸▸
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

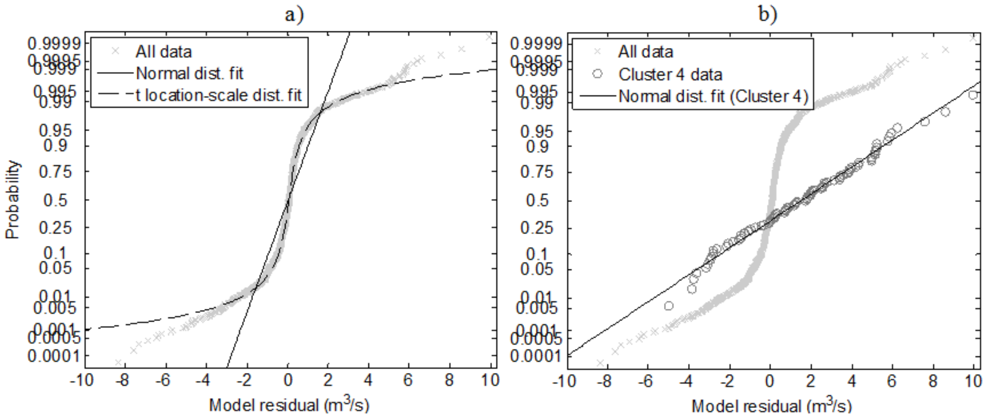

Figure 7. Probability plots for model residuals (during training) for Brue catchment: **(a)** comparison of the two fitted distributions: normal vs. *t* location-scale distribution, **(b)** the cluster representing high flow and high rainfall.

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

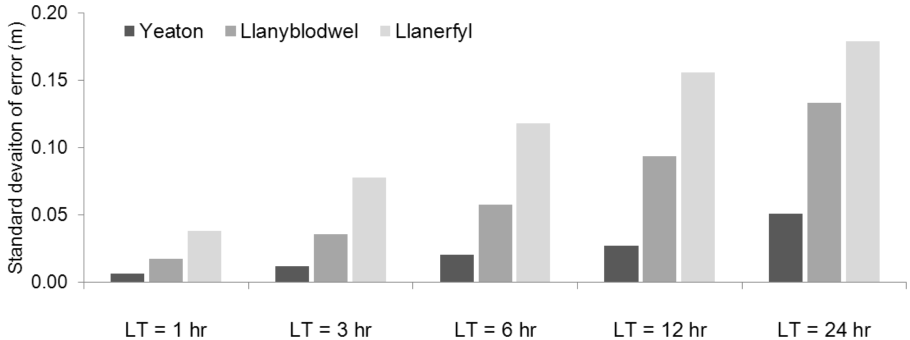

Figure 8. Standard deviation of model error (during calibration, Llanyblodwel).

HESSD
11, 10179–10233, 2014

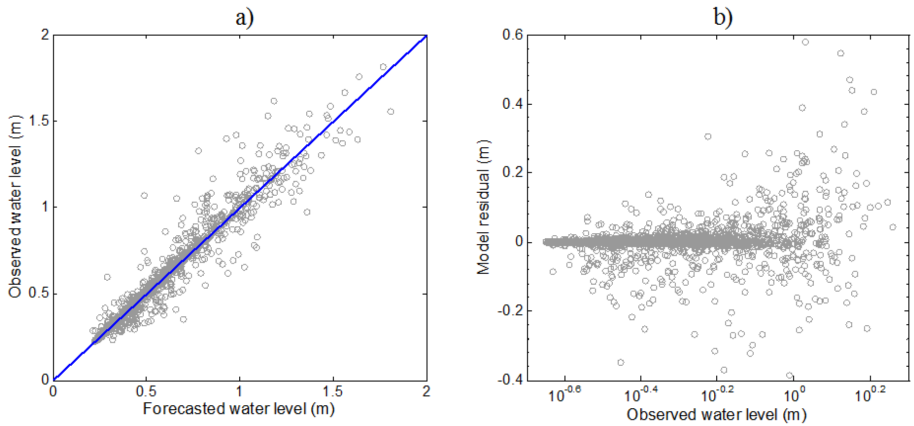

Figure 9. Observed water level, forecasted water level, and model residuals during calibration (Llanyblodwel, lead time = 6 h).

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

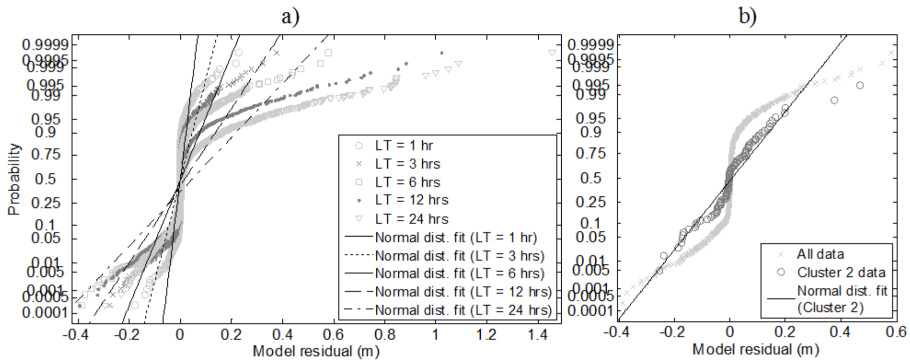

Figure 10. Probability plots for model residuals (during training) for Llanyblodwel catchment: **(a)** comparison of fitted normal distributions for all the lead times, **(b)** the cluster representing high groundwater level, high water level, and low soil moisture deficit (lead time = 6 h).

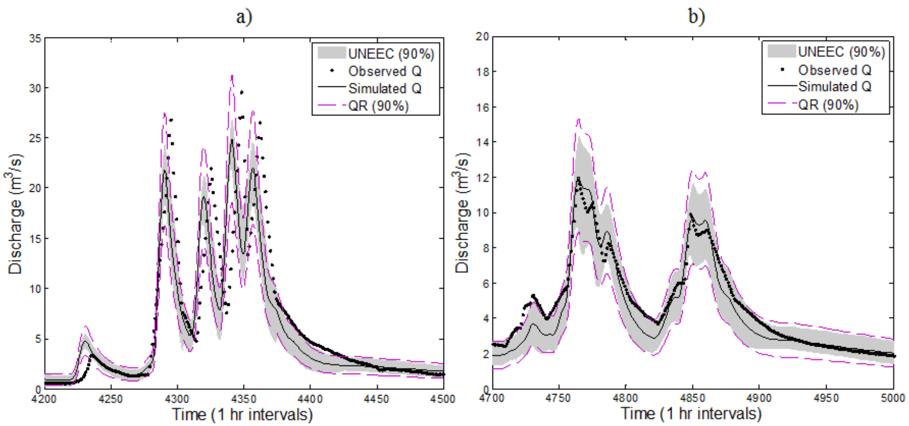

Figure 11. Comparison of prediction limits for 90 % confidence level during validation: **(a)** for the highest peak event (16 December 1995 04:00–28 December 1995 16:00), and **(b)** for a medium peak event (6 January 1996 00:00–18 January 1996 12:00).

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

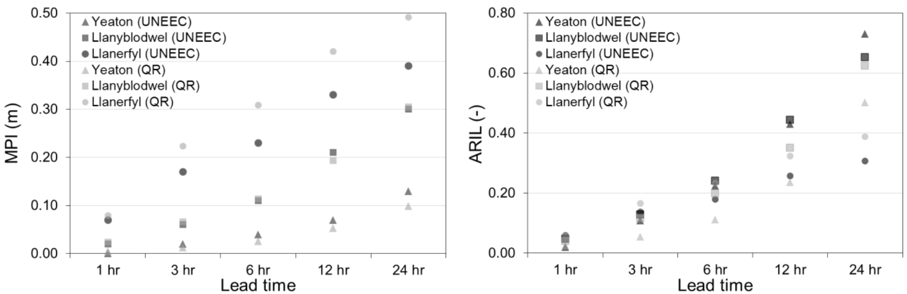

Figure 12. MPI (left) and ARIL (right) values obtained during validation period (7 March 2010 20:00–7 March 2013 08:00) for 90 % confidence level.

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

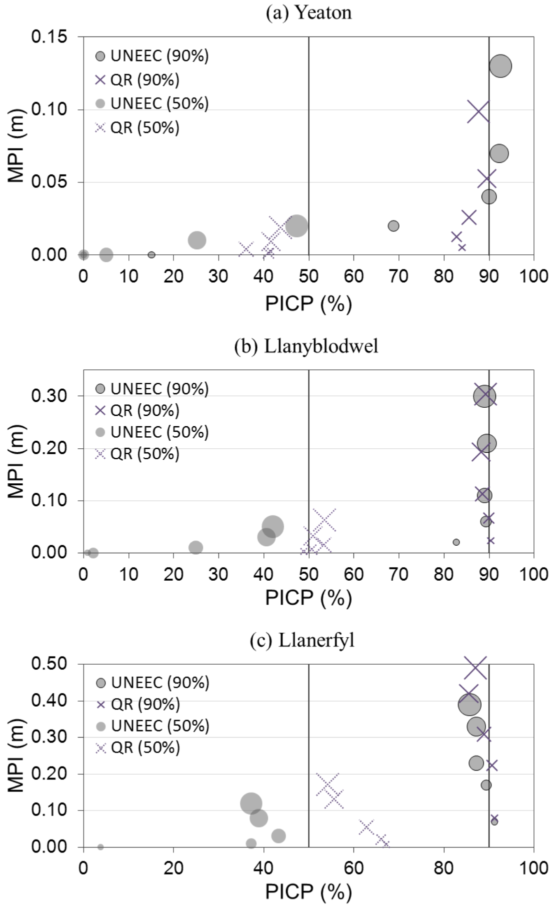

Figure 13. Comparison of UNEEC and QR based on both MPI and PICP during validation period (7 March 2010 20:00–7 March 2013 08:00) for 90 % confidence level (the size of the marker represents the lead time, i.e. bigger the marker, longer the lead time).

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

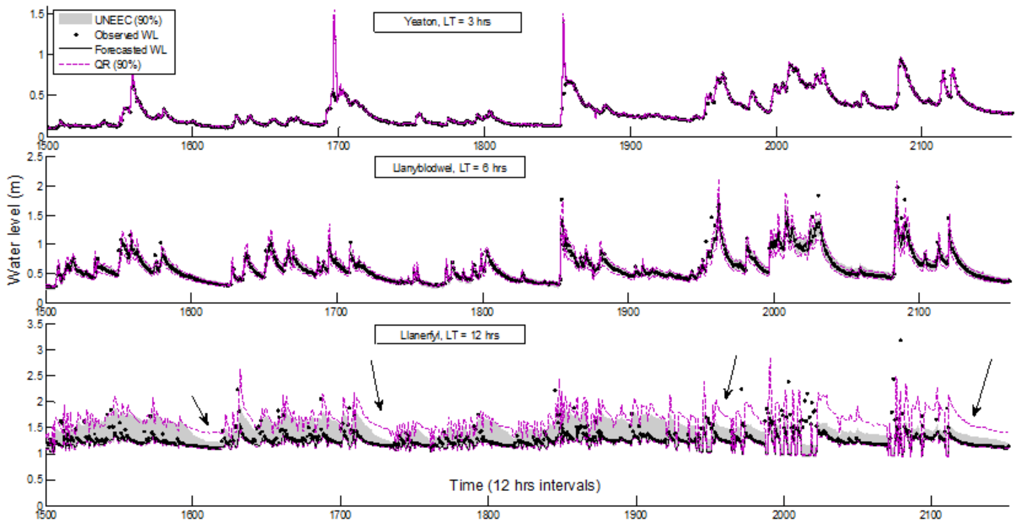

Figure 14. Comparison of prediction limits for 90 % confidence level during validation (1 April 2012–7 March 2013): **(a)** Yeaton, lead time = 3 h, **(b)** Llanyblodwel, lead time = 6 h, **(c)** Llanerfyl, lead time = 12 h.

HESSD
11, 10179–10233, 2014

Estimation of predictive hydrologic uncertainty using quantile regression

N. Dogulu et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion