Zurich, 27. June 2017

HESSD MS No.: hess-2017-75 – Authors’ Reply

Dear Editor,

Please find below our revised manuscript and the replies to the comments of reviewers.

We thank again all the parties involved in this review process.

Sincerely,

Omar

Omar Wani *(on behalf of the authors)*

***Please Note:** The reference of all line numbers refers to the revised manuscript without marked changes.*

Comment Reviewer 1 (J. Matos)

General comments
The manuscript proposes a non-parametric method to estimate the uncertainty associated with residuals of deterministic models applied to hydrologic forecasting.
The method relies on the well-known k-nearest neighbors (kNN) technique, being simple both conceptually and in its application. It is compared to two other post-processor techniques – Quantile Regression (QR) and Uncertainty Estimation based on local Errors and Clustering (UNEEC) – over two test cases in the UK. The paper is well written and clear. Also, the title and abstract fit the paper's contents. Most relevantly, in my opinion the paper addresses a relevant scientific question related to finding nonparametric methods of simple and broad application that allow accurate estimations of predictive uncertainty in hydrology, a question which falls into the scope of HESS. The authors evidence knowledge of the topic and relevant past publications. The analyses that were undertaken are well described, as is the kNN method that is proposed. The results of the comparisons with both QR and UNEEC are clear and obviously benefit from the work that some of the authors did and published on those models. I believe that the paper is a valuable contribution to the community and recommend its publication following the clarification of some aspects discussed in the specific comments.

The authors should adopt only one version of the terms "post-processor", "postprocessor", and "post processor".

> Thank you for pointing this out. We now use the term with consistency. "post-processor" when used as an adjective and "post processor" when used as a noun.

The provided URL: www.modeluncertainty.com does not seem to work (tested on two days roughly one week apart).

> We have fixed the script and removed the bugs. The website should be able to run for files with different input variable vector sizes and for different values of k. However, the format of the files that are uploaded should be the same as that of the example files.

The chosen measure of proximity was the Euclidean distance and the different variables on which the model is conditioned upon are normalized. Whereas in a model such as QR the importance of each variable is evaluated by the model, in kNN a variable's importance is defined a priori by its scale. The possibility of attributing different ways to input variables is mentioned, but not developed nor the target of a sensitivity analysis. I have doubts that kNN performs as well as QR (for example) if the informative inputs are mixed with less informative ones (and their weights are not properly adjusted). If so, the advantages in simplicity that kNN may have risk being offset by a more demanding input selection and preparation process.

> The reviewer makes an important point about the semi-quantitative guidelines that have been employed for the choice of the input variable vector for kNN resampling i.e correlation analysis and using simple input variable vectors. Whereas, UNEEC and QR use regression to assign proper weights to their

input variables. The authors acknowledge that kNN resampling will be contingent on the choice of the variable vector, however, we also contend that the simplest of input variable vector (conditioning only on simulated system response and error in the previous time step), would be able to produce acceptable performance in capturing uncertainty. We showed this by considering such an input variable vector for three subcatchments of Severn and ten different lead times (Figure 7 and Eq 15/16). However, as the concerns are well-founded, we now explicitly state this limitation in the abstract (Page 1/ Line 22). We have added further analysis for Brue catchment using three different input variable vectors and two values of k (Page 3/ Line 31, Page 12/ Line 21-24, Figure 9 and Eq 18, 19 and 20). We also mention these results in the discussion and conclusion part to state such a sensitivity.

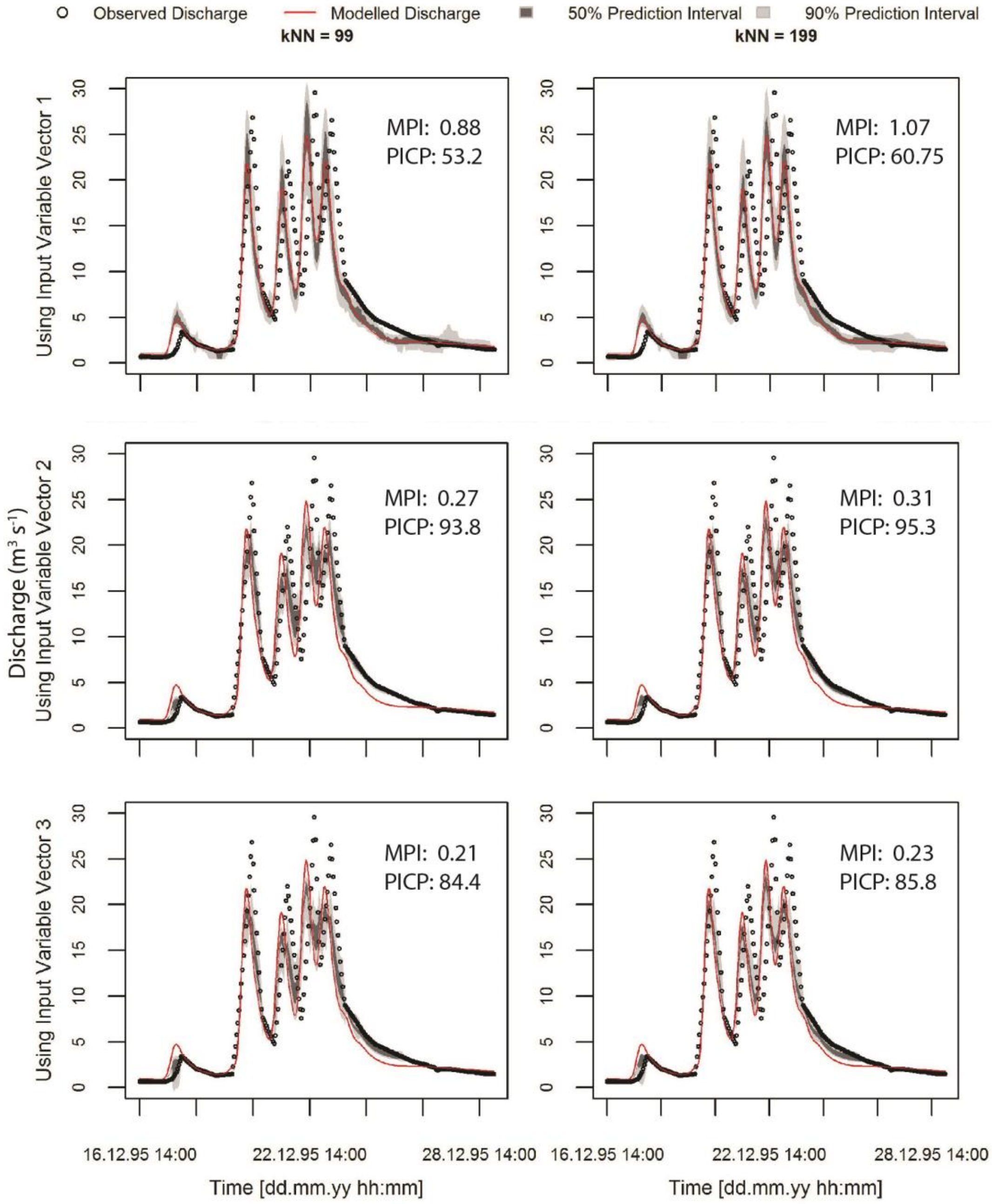


Figure 9. 50% and 90% prediction intervals for Brue catchment using kNN resampling. The hydrographs are shown for two different k values (99, 199) and three different input

variable vectors (Eq. (18), Eq. (19) and Eq. (20) for Input Variable Vector 1, 2 and 3 respectively). This is the largest event in the validation time series. (50% prediction interval is the interval between 25% and 75% quantiles of residual error, and 90% quantile is the interval between 5% and 95% quantiles. MPI and PICP correspond to whole validation time series.)

Line 339: Fig 5(a), the last sub-graph, the posterior distribution of the parameter ap Regarding validation metrics, it would be nice to add encompassing reliability assessment criteria such as the index α and the index ξ [e.g. Renard et al., 2010].

Thank you for this suggestion, we went through the suggested paper and have added Index Alpha to our analysis (As defined in Figure 3 of Renard et.al. 2010). We also feel that it is a valuable validation metric as it provided one number to capture the mismatch between the theoretical and observed quantiles of the error distribution. (We have added Eq 13 and Eq 14). The results are mentioned in a new Table 2.

We also went through the index ξ, and cite from the paper “Note that x = 1 does not imply perfect reliability. Consequently, this index is used primarily for detecting highly unreliable PDs.” Thus, this index is useful to capture the bad pdf at the extremes. As we do not used an explicit description of a likelihood function, the authors were not sure how to evaluate the graphical description of index ξ (Figure 3 from Renard et.al. 2010). Moreover, it comes out to be perfectly 1, for a quantile resolution of 1 percent, as used in this study. The authors also feel that the information that index ξ is supposed to convey is to a great extent conveyed by Figure 6 and 10 in our manuscript. We are hopeful that the reviewer will find our line of thought adequate as a reply.

It would be important to specify in the paper whether the inputs for kNN, UNEEC and QR are the same within each catchment. If this is not the case the authors should justify how may/do differences affect results and specify what part of the comparative performances is due to model structure (kNN, UNEEC and QR) and what part can be associated with the choice of input data.

Thank you for pointing this out. As far the choice of input vector is concerned, QR values for this study uses only the predicted system response (discharge/ water level) as the covariate for the quantile. As far as UNEEC is concerned, the they used correlation analysis to choose the input variable vector. As depicted in Dogulu et. al.2016, the configuration of QR and UNEEC was chosen based on some heuristic guidelines (average mutual information and correlation analysis) and the best possible option was chosen. In this paper our comparison is restricted to such configurations of QR and UNEEC. We stay vigilant not to claim a general verdict over performance of kNN resampling compared to all configurations of QR and UNEEC. However, we do think that the analysis for three subcatchments in Severn and one in Brue substantiates the proposition that kNN performance is comparable to QR and UNEEC.

For the Brue case study, apart from two other input variable vectors (Eq. 18/19), the same input variable vector (Eq. 20) was used as in the case of UNEEC.As mentioned earlier, the results are presented in Figure 9.

The Mean Prediction Interval (MPI) results presented in Table 2 are striking. I do not believe that the discussion on why the results obtained by kNN are so strikingly better than UNEEC's or QR's justifies such a remarkable improvement and would very much like to understand the underlying reasons. Related to the previous question, I would also like to see results on the performance of the kNN method when past error observations are not added to the input vector.

> We redid the analysis so that the MPI and PICP of the results can be calculated again on a less informative input variable vector, without the past errors in the input variable vector, and we got noticeable decrease in the performance of the technique. The new results are presented in Figure 9 and table 3.
>
> We now have elaborate on the small MPI by kNN resampling in the discussion.(Page 14/Line 18-24)

Please check the References for missing information (e.g. Sikorska et al.).

> Exact change made.

Technical comments

> Exact change made.

References

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, n/a-n/a, 10.1029/2009WR008328, 2010.

Dogulu, N., Lopez, P. L., Solomatine, D. P., Weerts, A. H., and Shrestha, D. L.: Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments, Hydrol Earth Syst Sc, 19, 3181-3201, 10.5194/hess-19-3181-2015, 2015.

Comment Reviewer 2 (L. Raso)

The manuscript explores and discusses the application of k-Nearest Neighbors (kNN) method, a non-parametric machine learning technique, to estimate the predictive uncertainty in heteroschedastic streamflow forecasting. The paper is clearly written. It comes completed of a internet website where a user friendly interface makes application of kNN straightforward. The innovation is well framed in the recent literature on predictive uncertainty of heteroschedastic processes in hydrology, giving particular attention to comparable methods that estimates predictive uncertainty a posteriori. The authors clearly present advantages and limits of kNN with respect to other methods. Nonetheless, as already mentioned by the other referee, the limits of kNN in extrapolation, presently mentioned only in the conclusion, should deserve more emphasis. The manuscript brings a valid and innovative contribution to its field, and I suggest its acceptance. There are two issues, however, that could contribute to make the case for this methodology in a more convincing way, and some minor issues that deserve at least to be mentioned.

The first main issue regards the selection of the k value, i.e. the number of data points considered similar to the instance to be estimated. Fixing k is a problem of kNN method. In general, when kNN is used for prediction, k is selected in order to maximize the predictive capacity, tested by a cross-validation on data. In the manuscript the criteria for selecting k is the stabilization of residuals probability distribution. Change in residuals distribution is quantified by the cumulative difference, defined at Equation (17). The reason why the stabilization of residuals distribution is a good criteria for fixing k is not clear. Moreover, this value is monotonic, hence it does not offer a clear-cut rule. The authors propose that k is to be selected when shape changes, but this rule, differently from what stated ad page 6 line 8, is not fitted to be used in an optimisation procedure. The second issue regards the estimation of quantiles. kNN use the closest k values to build up an empirical distribution made of situations (i.e. data-points) similar to the “true” distribution that one intends to estimate. When kNN is applied for regression, the value to be predicted is the expected value, then the algorithm takes the average of k nearest data-points. In the proposed application instead, the empirical distribution is used to estimate some quantiles. Quantile estimation, however, has a different convergence rule than the expected value, particularly critical in estimating tails. Convergence rules of empirical distribution at quantiles of interest is well described in [1], chapter 21. Error in quantile estimation decreases with root square of k, and it is larger for quantiles close to 1 and 0. Using the 99th value from a set of 100 points as estimator of the 99th quantile may not be sufficient in guaranteeing sufficient convergence. Quantile estimation from empirical distribution introduce an error that must be considered, or at least discussed.

> The reviewer puts forth an important aspect of sampling errors due to the finitude of samples - the first two moments tend to be less prone to errors compared to the tail estimates of a distribution. As in this study we use on 99 samples to generate quantiles corresponding to 5% and 95%, the errors are not capped by $O(1/(k)^{0.5})$. Therefore, now the sensitivity of the technique to the sample size is didactically shown using two values of k (99 and 199) - its impact on PICP and MPI for Brue catchment. We have extended Figure 9. Also, we now mention this dependence on k early on in the manuscript. (page 5/Line 24-29, page 6 line 1-2). We thank the reviewer for bringing our attention to an informative piece of literature on convergence statistics (van der Vaart, 1998). We have also added it as a reference in the manuscript.

“Moreover, this value is monotonic, hence it does not offer a clear-cut rule. The authors propose that k is to be selected when shape changes, but this rule, differently from what stated ad page 6 line 8, is not fitted to be used in an optimisation procedure.”

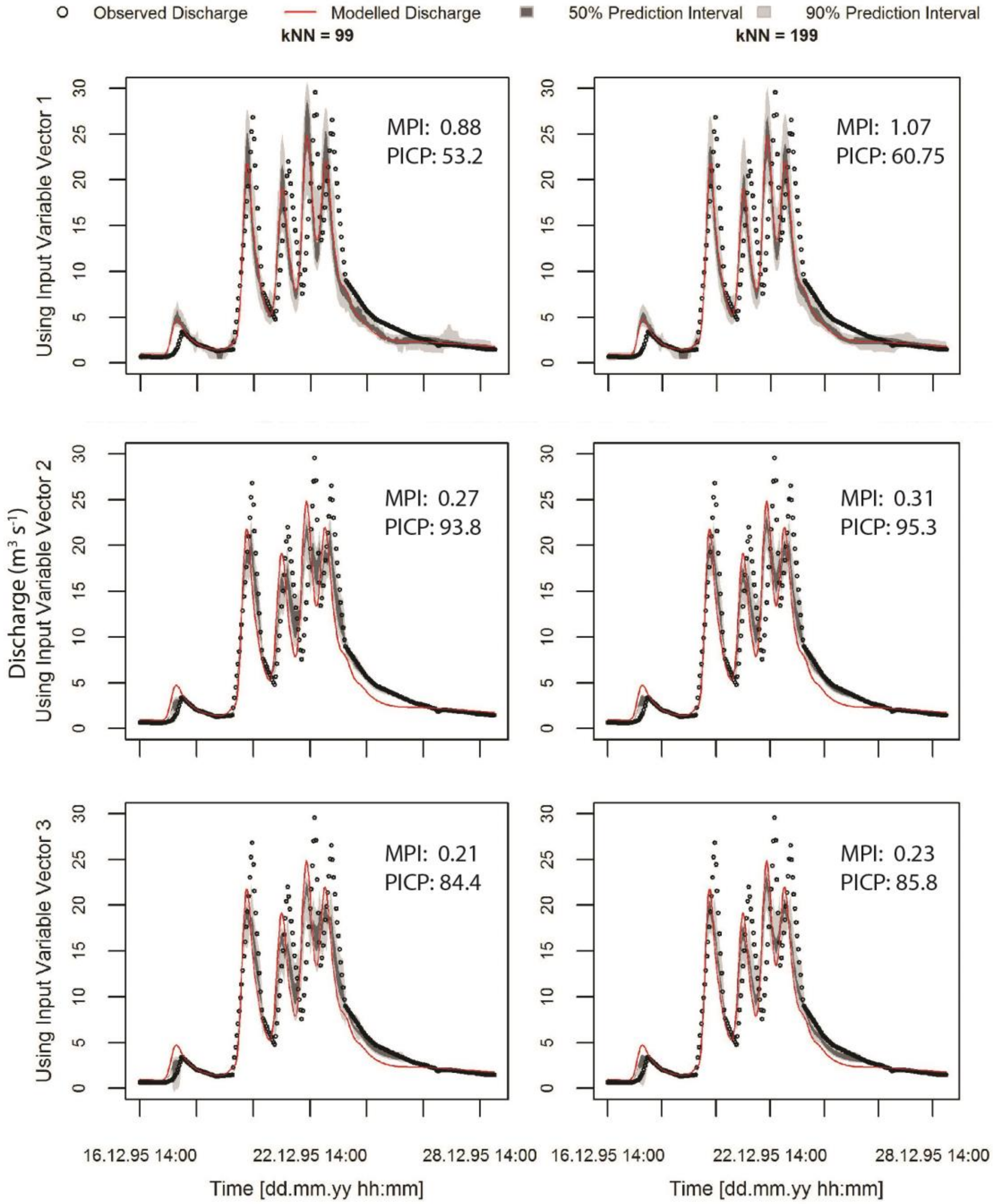


Figure 9. 50% and 90% prediction intervals for Brue catchment using kNN resampling. The hydrographs are shown for two different k values (99, 199) and three different input variable vectors (Eq. (18), Eq. (19) and Eq. (20) for Input Variable Vector 1, 2 and 3 respectively). This is the largest event in the validation time series. (50% prediction interval is the interval between 25% and 75% quantiles of residual error, and 90% quantile is the interval between 5% and 95% quantiles. MPI and PICP correspond to whole validation time series.)

We agree that Eq. 17 is a synthetic index and does not capture many aspects of distributional convergence. Nonetheless, the authors used it as a simple heuristic tool to agree on a reasonable value of k. The monotonicity of this index captures two aspects of the changing distribution. For small values of k, the “Cumulative Difference” changes a lot and then the sampling error decreases. However, to incorporate the concerns of the reviewer, we have removed the line referring to optimization on k Page 6, Line 16, which could have been misconstrued as an optimization exercise for k value carried out in this study. Also, as mentioned before, we have added analysis related to the sensitivity of k (Figure 9). We don’t seem to notice worrying changes in the PICP and MPI of the Brue catchment when changing k from 99 to 199. However, as expected, the MPI does get somewhat bigger. We discuss this dependence explicitly on Page 14/Line 20-24.

In the discussion on verification index, the authors show that they are aware of thelimits in using few indicators. The authors state that "PICP and MPI [...] give a reasonable assessment of performance". But this is not further explained. There are likely good reasons to select these indicators, but this should be better explained in the text, considering also that the application is about flood forecasting.

The idea of checking PICP at 90 uncertainty bands is a common practice in hydrology. However, more tail events might become interesting for design problems, then the reliability of the whole error distribution is more interesting than the mere computation of PICP and MPI. Taking the advice of reviewer 1, we have also added another metric for performance – the Index Alpha (Eq. 13/14). And the results are presented in Table 2.

In Equation 7, variables are standardized one at a time, losing information about covariance. Why not considering variables as a multidimensional distribution, then using the covariance matrix to standardise? This would make use of the mutual information about variables in a more efficient way.

The individual normalization prevents the kNN conditioning to be exclusive to the dimension with higher variance. However, as the reviewer points out, there can be more advanced ways of normalization, like the usage of covariance matric which captures the linear dependence between different dimension of the input variable vector. We have not used tested techniques in this research. We are aware that better metrics to choose of input variable vector and for the normalization can improve the technique significantly. Hopefully in the future we can carry out more analysis to better delineate such technique.

Other comments

Page 3, line 24: add "than" after simpler

Exact change made.

Page 5, line 12: "uncertainty in observational data is not considered", why can not it be included?

Exact change made.

Page 8 line 23: remove extra dot.

Exact change made.

Page 9 line 17: the adverb "just" looks like non necessary.
Exact change made.

Page 10 line 11: The result description would be easier to follow if the reference to the figure was placed at the beginning of this paragraph (from line 19 to line 11).
Exact change made.

References

van der Vaart, a W.: Asymptotic Statistics, Asymptot. Stat., 3, 443, doi:10.2307/2530729, 1998.

Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting

Omar Wani[1,2,a], Joost V.L. Beckers[2], Albrecht H. Weerts[2,3], Dimitri P. Solomatine[1,4]

[1]UNESCO-IHE Institute for Water Education, Delft, the Netherlands
[2]Deltares, Delft, the Netherlands
[3]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, Wageningen, the Netherlands
[4]Water Resource Section, Delft University of Technology, Delft, the Netherlands
[a]currently at: Institute of Environmental Engineering, Swiss Federal Institute of Technology (ETH), Zürich and Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland

Correspondence to: O. Wani (owani@student.ethz.ch)

Abstract

A non-parametric method is applied to quantify residual uncertainty in hydrologic streamflow forecasting. This method acts as a post ~~-~~processor on deterministic model forecasts and generates a residual uncertainty distribution. Based on instance-based learning, it uses a k-nearest neighbour search for similar historical hydrometeorological conditions to determine uncertainty intervals from a set of historical errors, i.e. discrepancies between past forecast and observation. The performance of this method is assessed using test cases of hydrologic forecasting in two UK rivers: Severn and Brue. Forecasts in retrospect were made and their uncertainties were estimated using kNN resampling and two alternative uncertainty estimators: Quantile Regression (QR) and Uncertainty Estimation based on local Errors and Clustering (UNEEC). Results show that kNN uncertainty estimation produces accurate and narrow uncertainty intervals with good probability coverage. Analysis also shows that the performance of this technique depends of the choice of search space. Nevertheless, ~~T~~the accuracy and reliability of ~~these~~ uncertainty intervals generated using kNN resampling are at least comparable to those produced by QR and UNEEC. It is concluded that kNN uncertainty estimation is an interesting alternative to other post processors, like QR and UNEEC, for estimating forecast uncertainty. ~~An advantage of this method is that~~Apart from ~~-~~its concept being~~is~~ simple and well understood, an advantage of this method is that it is relatively easy to implement.~~ and it requires little tuning.~~

1 Introduction

Hydrologic forecasts for real-life systems are inevitably uncertain (Beven and Binley, 1992; Gupta et al., 1998; Refsgaard et al., 2007). This, among other things, is due to the uncertainties in the meteorological forcing, in the modelling of the hydrologic system response and in the initial state of the system at the time of forecast. It is well accepted that, compared to a simple deterministic forecast, additional information about the expected degree of accuracy of that forecast is valuable and generally leads to better decision making (Krzysztofowicz, 2001). Various techniques have therefore been developed to quantify uncertainties associated with the meteorological model input (van Andel et al., 2013), the initial state of the model (Li et al., 2009) and with the hydrologic models themselves (Deletic et al., 2012; Coccia and Todini, 2011). Frameworks and guidelines have been developed to incorporate uncertainty analysis of environmental models effectively in decision making (Arnal et al., 2016; Reichert et al., 2007; Refsgaard et al., 2007). Broadly, there are three basic approaches to uncertainty estimation: i) explicitly defining a probability model for the system response e.g. (Todini, 2008), ii) estimation of statistical properties of the error time series in the post-processing phase of model forecast e.g. (Dogulu et al., 2015) and iii) methods using Monte Carlo sampling of inputs and/or parameters, aimed~~ed~~ at getting a range of ~~quantifying the output probabilistically~~model outputs e.g. (Beven and Binley, 1992; Freer et al., 1996). ~~Many~~ Other uncertainty estimations techniques may employ a combination of these approaches (~~Montanari and Brath, 2004;~~ Del Giudice et al., 2013). Some techniques focus on one source of uncertainty, such as the model parameter uncertainty (Benke et al., 2008) or the model structure uncertainty (Butts et al., 2004), while others focus on combined uncertainties stemming from model parameters, model structure deficits and inputs (Schoups and Vrugt, 2010; Evin et al., 2013; Del Giudice et al., 2013). In this context, it is important to note that apart from estimating uncertainty of model parameters during calibration, uncertainty estimation for hydrologic forecasting requires quantification of predictive uncertainty, which ~~includes~~ includes uncertain system response additionally to ~~new~~ different combinations of model ~~input~~ parameters (Renard et al., 2010; Coccia and Todini, 2011; Dotto et al., 2012).

In this paper, we will restrict ourselves to the class of uncertainty estimators called post ~~-~~processors. These methods usually do not discriminate between different sources of uncertainty. They “aggregate” all sources into a so-called residual uncertainty. Post-processing methods assume the existence of a single calibrated model with an optimal set of model parameters, and build a statistical or machine learning model of the residual uncertainty. Typically, these techniques relate a combination of model inputs and/or outputs to the model error distribution. Various post ~~-~~processors have been developed and applied to hydrologic modelling, such as a meta-Gaussian error model (Montanari and Brath, 2004), ~~Quantile Regression (Weerts et al., 2011),~~ UNEEC (Solomatine and Shrestha, 2009), Quantile Regression (Weerts et al., 2011), and DUMBRAE (Pianosi and Raso, 2012). Quantile Regression (QR) is a relatively straightforward post-processing technique that relates the probability of residual errors to the model forecast (the predictand) by a regression model that is derived from historical forecasts and observations. QR has been successfully applied for uncertainty quantification in hydrologic forecasts with various modifications (Weerts et al., 2011; Verkade et al., 2013; Roscoe et al., 2012; López López et al., 2014; Hoss and Fischbeck, 2015). Whereas, UNEEC involves a machine learning technique for building a non-linear regression model of error quantiles (Solomatine and Shrestha, 2009). UNEEC includes three steps: 1) Fuzzy clustering of input data in the space of “relevant” variables; 2) Estimating the probability distribution function of residual errors for each cluster and 3) building a machine learning model (e.g. an artificial neural network) of the prediction interval for a given probability

(Dogulu et al., 2015). Many other uncertainty estimation techniques, such as DUMBRAE (Pianosi and Raso, 2012), HUP (Krzysztofowicz, 1999), Model Conditional Processor (Coccia and Todini, 2011), Bayesian revision (Reggiani et al., 2009) and Bayesian Model Averaging (Raftery et al., 2005) make explicit assumptions about the nature of the probability distribution function of error. This is not necessary for QR and UNEEC (Lopez et al., 2014; Dogulu et al., 2015). Nevertheless, in QR and UNEEC assumptions need to be made about the form of the regression function that is used to calculate the quantiles.

In an attempt to explore the utility of easier-to-implement post-processing techniques, we employ a simple non-parametric forecast method for residual uncertainty quantification. This method uses kNN search to learn about the past residual errors, which avoids having to make explicit assumptions about the nature of the error distribution and tuning of distribution parameters. Instance-based learning has been used in meteorology and hydrology before for resampling of precipitation and streamflows, most notably by Lall and Sharma (1996), who used the k-nearest neighbour (kNN) method for resampling of monthly streamflow sequences. kNN search has also been used in a non-parametric simulation method to generate random sequences of daily weather variables (Rajagopalan and Lall, 1999). They defined a weighting function for probability where the predictand is resampled from k values. Jules and Buishand (2003) used nearest-neighbour resampling to generate multi-site sequences of daily precipitation and temperature in the Rhine basin. Also, instance-based learning has been used as a data-driven model for hydrologic forecasting (Solomatine et al., 2008;Solomatine and Ostfeld, 2008). Beckers et al. (2016) use nearest neighbour resampling to generate monthly sequences of climate indices and related precipitation and temperature series for the Columbia River basin. Specifically in the context of error modelling, a version of UNEEC that uses kNN instance-based learning as its basic machine learning technique to predict the residual error quantiles, was compared to the original ANN-based UNEEC in Shrestha and Solomatine (2008). However, kNN can be also used without the complicated UNEEC procedure that includes fuzzy clustering. The application of kNN has ~~also~~ recently been tested for forecast updating by constructing a deterministic error prediction model (Akbari and Afshar, 2014). Similarly, it has been shown that model errors can be resampled using kNN, after explicitly accounting for input and parameter uncertainty, to generate uncertainty intervals (Sikorska et al., 2015). In this paper we extend the simplification of kNN resampling for uncertainty estimation. We present an application of the kNN method to generate residual uncertainty estimates for a predictand, using a fixed time series of input and fixed model parameters, and explore if this approach, being simpler that many other uncertainty quantification approaches mentioned above, is a useful or even a better alternative.

To demonstrate its use, we employ a relatively simple configuration of kNN resampling to produce uncertainty intervals for hydrologic forecasting. The next section explains the method in more detail and describes the validation procedure, i.e. the performance indicators. I~~Next, i~~n section 3, the method is applied to two case studies, each with a different system response (discharge and water level). The performance of kNN uncertainty estimation as a function of forecast lead time is analysed in the first case study. Second case study is used to further validate the performance of kNN uncertainty estimation and analyse its sensitivity to the choice of search space and the value of k. Also, the influence of systematic bias in the hydrologic model on the uncertainty intervals generated by kNN search is explored in the second case study. For both case studies, performance indices of kNN resampling are compared to those of QR and UNEEC. And finally in section 4, we discuss the usability of kNN search as a post-processor uncertainty estimator in hydrologic forecasting.

2 Method

2.1 kNN error model

The kNN residual uncertainty estimator can be seen as a zero$^{\mathrm{th}}$ order local error quantile model built from a kNN search. Let us define a vector $\boldsymbol{v}$ in n-dimensional space of variables (the search space) on which the residual uncertainty is assumed to be statistically dependent.

$$\boldsymbol{v} = [v^1, \dots, v^{\mathrm{n}}] \tag{1}$$

The cumulative probability distribution function C of residual errors at prediction time-step t conditioned on $\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}$ is defined as:

Formatted: Font: Italic

$$C_{\mathrm{t}}(e|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) = P_{\mathrm{t}}(E \le e|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) \tag{2}$$

Where P is the probability function and E denotes the random variable for residual errors. Residual error is defined throughout this paper as the difference between the simulated values and the observed values for a hydrologic ~~quantity~~ system response f, like discharge or water level.

$$e = f_{\mathrm{simluated}} - f_{\mathrm{observed}} \tag{3}$$

We are making the assumption of stationarity in time so that past error distributions are representative of the future:

$$C_{\mathrm{t}}(e|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) = C_{\mathrm{p}}(e|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) \tag{4}$$

The subscript p denotes historical time series. Therefore C_{p} is the cumulative distribution function of residual errors from the past. In Eq. (4), C_{p} is being conditioned to the input variable vector at time t. Nevertheless, as we only have single realizations of the error variable E for each historical point, we relax the constraint of $\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}$. Instead, we assume that the nearby neighbours of $\boldsymbol{v}_{\mathrm{t}}$ in n-dimensional space will have a similar probability distribution of errors as $\boldsymbol{v}_{\mathrm{t}}$ and that these historical errors are samples from $C_{\mathrm{p}}(e|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}})$**.** An empirical probability distribution can thus be constructed using the kNN historical errors:

$$C_{\mathrm{t}}(e_{\mathrm{t}}|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) \approx C_{\mathrm{p}}(e|\boldsymbol{r}_{\mathrm{p}} \le \boldsymbol{r}_{\mathrm{k}}) \tag{5}$$

where $\boldsymbol{r}_{\mathrm{p}}$ is the Euclidean distance in n-dimensional space of input variables.

$$\boldsymbol{r}_{\mathrm{p}} = |\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{\mathrm{t}}| = \sqrt{\left[\sum_{\mathrm{i}=1}^{\mathrm{n}} (v_{\mathrm{p}}^{\mathrm{i}} - v_{\mathrm{t}}^{\mathrm{i}})^2\right]} \qquad (6)$$

$\boldsymbol{v}_{\mathrm{p}}$ is the input variable vector of the past data point in the cloud of such past data points $\boldsymbol{v}$ (Figure 1)and $\boldsymbol{r}_{\mathrm{k}}$ is the distance to the k$^{\mathrm{th}}$ nearest neighbour of $\boldsymbol{v}_{\mathrm{t}}$. Choice of the input variable vector is a problem in itself since it should include only the most relevant variables that determine the forecast uncertainty. In this study, the input variable vector is chosen based on correlation between the candidate variables and the past errors. If the correlation between the error time series and a particular candidate variable is relatively high, then it can be included in the input variable vector space. Other, more sophisticated methods involving the mutual information can be used as well (Fernando et al., 2009). This will be exemplified in the case studies described in the next section. To represent relative importance of input variables used in the search, dimensions of the input variable vector space can be suitably weighted in. Also, the model-based methods can be used where~~n~~ ~~the~~ models are built for each considered candidate input variable~~s~~ set ~~considered~~ and the choice is made based on their relative performance. These, however, were not explored in this study; it rather focused ~~only~~ on the usability of ~~k nearest neighbour~~kNN search in its most basic implementation for uncertainty quantification. Nevertheless, we do demonstrate the sensitivity of the uncertainty intervals on the choice of input variable vector.

In order to~~To~~ level variables with different magnitudes, they are normalized. If σ_{i} represents the standard deviation of input variable i~~I~~ calculated using the past data, then:

$$\boldsymbol{r}_{\mathrm{p}} = \sqrt{\left[\sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(v_{\mathrm{p}}^{\mathrm{i}} - v_{\mathrm{t}}^{\mathrm{i}})^2}{\sigma_{\mathrm{i}}^2}\right]} \qquad (7)$$

Once, the input variable vector space is decided, the probability of non-exceedance of a forecast error is calculated empirically by sampling from the conditional error distribution:

$$C_{\mathrm{t}}(e_{\mathrm{t}}|\boldsymbol{v} = \boldsymbol{v}_{\mathrm{t}}) \approx C_{\mathrm{p}}(e|\boldsymbol{r}_{\mathrm{p}} \leq \boldsymbol{r}_{\mathrm{k}}) = j/k \qquad (8)$$

where j is the rank of value e (for which the probability of non-exceedance is being computed) in the ascending array of k error values. The kNN search is thus employed to generate a sample and to build an empirical error distribution for this predictive uncertainty quantification. Such a mathematical description does not employ explicit regression models for predicting quantiles, which can be seen as a disadvantage in extrapolating outside available data. Also, ~~A~~as th~~is~~ise configuration of kNN used in this research generates residual error quantiles, which capture the mismatch between measurement values and simulated values, the uncertainty in observational data is not considered. The generated quantiles are aimed to capture the measured system response and do not attempt to capture the true response of the hydrologic system.

As one would expect, due to the nature of our sampling approximation (Eq. (8)), the number of nearest neighbours, k, will affect the empirical conditional probability distribution of errors. If k is very large, many data points that are quite distant from $\boldsymbol{v}_{\mathrm{t}}$ (Figure 1) will be selected and the conditioning on the current forecast situation will not be valid. Large values of k will thus yield error distributions with larger uncertainty intervals - resembling the marginal error distribution. If k is small, the set of k errors will be small and subject to sampling error, so this set will not adequately represent the uncertainty distribution at $\boldsymbol{v}_{\mathrm{t}}$. The tail of a distribution is more prone to sampling errors compared to its mean. Thus, to attain an acceptable degree of convergence, many more samples are required for quantiles corresponding to bigger prediction intervals (van der Vaart, 1998). For improved performance, the value of k can be subject to optimisation of some cost function: the optimal value of k could be the one that enables a reasonable estimate of the uncertainty quantiles and additionally we may require that the sensitivity of the error distribution to k is small. In this study, we carry out such optimization using quite a simple heuristic guideline - the value of k is varied until the probability distribution of errors stabilizes and becomes less sensitive on the value of k for a few model predictions. ~~This will be~~ We also demonstrate the sensitivity of uncertainty intervals to the value of k ~~d by an example~~ in one of the case studies. The choice of this relatively simple procedure for error quantile generation using kNN resampling is a reasonable starting point to assess its potential for residual uncertainty. This study explores the potential of uncertainty estimation using kNN in as simple a way as possible. And then compare its performance to two other residual uncertainty estimators. More advanced application of kNN, for example using fuzzy weights and kNN sampling to assign prediction intervals (Shrestha and Solomatine, 2008) or through explicit consideration of uncertainty in parameter and input by sampling them from their distributions, has been successfully shown (Sikorska et al., 2015).

To summarize, the steps for uncertainty quantification using kNN resampling are as follows:

1. Compose the input variable vector space ($\boldsymbol{v}$) on which uncertainty will be conditioned. Correlation analysis can help find the most relevant variables.
2. Set the number of neighbours k.~~ (It can be identified by optimization as well).~~
3. For a forecast at prediction time-step t, identify the set of k nearest neighbours to the input vector $\boldsymbol{v}_{\mathrm{t}}$. This set represents the hindcasts (forecasts in retrospect) most similar to $\boldsymbol{v}_{\mathrm{t}}$.
4. Use the residual errors from these k points to build an empirical error distribution for the forecast at time-step t.
5. Finally, identify the errors corresponding to the required quantiles (probabilities of non-exceedance) from this empirical distribution (In this paper, we use 5-95% and 25-75% quantiles).

2.2 Validation methods

~~Two~~ Three statistical measures have been employed in this study to check the effectiveness of uncertainty estimation techniques, namely Prediction Interval Coverage Probability ($\mathrm{PICP_{PI}}$) ~~and~~ , the Mean Prediction Interval ($\mathrm{MPI_{PI}}$) (see, e.g.

Shrestha and Solomatine 2008; Dogulu et al., 2015) and Index Alpha (Renard et al., 2010). $\mathrm{PICP}_{\mathrm{PI}}$ represents the percentage of observations (*C*) covered by a prediction interval (PI) corresponding to a certain probability of occurrence (in our case 90% and 50%).

$$\mathrm{PICP}_{\mathrm{PI}} = \frac{N_{\mathrm{in}}}{N_{\mathrm{obs}}} \times 100\% \tag{9}$$

where N_{in} is the number of observations located within the PI and N_{obs} is the total number of observations. These metrics are calculated using the following equations:

$$\mathrm{PICP}_{90} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} C_{90} \times 100\%,\quad \mathrm{PICP}_{50} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} C_{50} \times 100\% \tag{10}$$

$$C_{90} = \begin{Bmatrix} 1, & \text{if } q_{\mathrm{i},0.05} \le q_{\mathrm{i}} \le q_{\mathrm{i},0.95} \\ 0, & \text{else} \end{Bmatrix},\; C_{50} = \begin{Bmatrix} 1, & \text{if } q_{\mathrm{i},0.25} \le q_{\mathrm{i}} \le q_{\mathrm{i},0.75} \\ 0, & \text{else} \end{Bmatrix} \tag{11}$$

~~$\mathrm{PICP}_{50} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} C_{50} \times 100\%$~~ ~~(12)~~

~~$C_{50} = \begin{Bmatrix} 1, & \text{if } q_{\mathrm{i},0.25} \le q_{\mathrm{i}} \le q_{\mathrm{i},0.75} \\ 0, & \text{else} \end{Bmatrix}$~~ ~~(13)~~

where $\mathrm{q}_{\mathrm{i},0.95}$ and $\mathrm{q}_{\mathrm{i},0.05}$ are values with 95% and 5% probability of non-exceedance at time i. Thus the region bound within these two values will have a confidence interval of 90%. Similarly, $q_{\mathrm{i},0.75}$ and $q_{\mathrm{i},0.25}$ represent the boundaries for 50% *C*. The MPI is the average width of the confidence intervals corresponding to a particular probability. It is a measure of the magnitude of the uncertainty.

$$\mathrm{MPI}_{90} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} (q_{\mathrm{i},0.95} - q_{\mathrm{i},0.05}),\qquad \mathrm{MPI}_{50} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} (q_{\mathrm{i},0.75} - q_{\mathrm{i},0.25})$$

(12~~4~~)

We also quantify the reliability of the predicted error quantiles by comparing it to the observed error quantiles. The mismatch between the observed ($q_{\mathrm{obs,j}}$) and predicted ($j/100$) error quantiles can be summarized by the Index Alpha (α).

$$\alpha' = \frac{1}{100}\sum_{\mathrm{j}=1}^{100} |\, q_{\mathrm{obs,j}} - j/100| \tag{13}$$

$$\alpha = 1 - 2\alpha' \tag{14}$$

There have been discussions whether an isolated verification index can capture all the aspects that make a probabilistic forecast good or bad (Laio and Tamea, 2007). The choice of a verification index for an uncertainty estimation technique should also be dependent on the purpose of hydrologic forecast. For example, Coccia and Todini (2011) evaluate the performance of Model Conditional Processor for flood forecasting using the predicted and observed probability of exceedance over a threshold. Also, in their study predicted error quantiles are compared to observed error quantiles. López López et al. (2014) and Dogulu et al. (2015) use PICP and MPI, among other verification measures, to access the performance of QR and UNEEC. This study will limit the comparison of kNN resampling with other techniques to PICP and MPI only, which give a reasonable assessment of performance. Nevertheless, it does not preclude the possibility that the uncertainty estimation techniques perform differently if evaluated using other indices.

3 Case studies

The performance of kNN resampling was evaluated by applying the technique to hydrological forecasting for several catchments in two different parts of England. The two case studies provide two different hydrologic conditions for testing and include different models for prediction. Also, different kinds of system responses are being predicted in the two case studies – water level and discharge. The accuracy of the quantified prediction intervals was deduced by using validation data sets. Also, the first case study was used to evaluate the impact of changing lead time on uncertainty of hydrologic models and its quantification using kNN resampling.

3.1 Upper Severn catchment

3.1.1 Catchment description

Upper Severn region is located in the Midlands, UK (Figure 2). River Severn, with a total length of 354 km, is the longest river in the UK. Its course acts as a geographic delineation between England and Wales, finally draining into the Bristol Channel. The overall River Severn catchment area is 10,459 km^2. Around 2.3 million people live in this region. The area is predominantly rural, but there are also a number of highly urbanized parts. The area covering the upper reaches of River Severn, from its source on Plynlimon to its confluence with the River Perry upstream of Shrewsbury in Shropshire, is called the Upper Severn catchment. The Upper Severn catchment is predominantly hilly. It is dominated on the western edge by the Cambrian Mountains and a section of the Snowdonia National Park (River Severn CFMP. EA, 2009).

The Severn catchment has a diverse geology. The headwaters of the ~~River Severn~~river rise on Silurian mudstones, siltstones and grits and flow eastwards over these same rock formations. These rock formations do not allow water to flow easily through them. Therefore they are classified as non-aquifers with only limited potential for groundwater abstraction. Further west, in the Middle Severn section, the River Severn encounters sandstones, which are classified as a major aquifer and are highly permeable, highly productive and able to support large groundwater abstractions (River Severn CFMP. EA, 2009). The climate of the Severn catchment is generally temperate, experiencing modest to high precipitation depending~~ent~~ on topography. Welsh Mountains can receive over 2,500 mm of precipitation per annum, whereas the rest of the catchment receives rainfall similar to the UK average - less than 700 mm per annum. The test forecast locations used in this study are Llanerfyl, Llanyblodwel and Yeaton. Table 1 lists the basin and hydrological information for these subcatchments (López López et al., 2014).

3.1.2 Experimental setup

Flood forecasting system for River Severn is organized in a sequential manner, being composed of a number of separate systems that are effectively linked. This forecasting system works with a high degree of automation and efforts have been made to involve a minimum amount of human intervention.~~ .~~ UK Environment Agency uses Midlands Flood Forecasting System (MFFS) to do flood forecasting and to help in warning operation. MFFS in turn is based on Delft-FEWS, Flood Early Warning System) platform (Werner et al., 2013). Within MFFS, there are lumped numerical models for rainfall-runoff (MCRM; Bailey and Dobson, 1981) and models for hydrologic (DODO; Wallingford, 1994) and hydrodynamic routing (ISIS; Wallingford, 1997). The rainfall input for MFFS is acquired from ground measurements via rain gauges, from radar measurements or from numerical weather prediction data. MFFS predict ahead in time the response of the Upper Severn subcatchments but, as expected, the quality of forecast deteriorates with increasing lead time.

To do uncertainty analysis for MFFS, hindcasting or reforecasting is done and then ~~simulations~~ results are compared to the observed data. All the input time series used for hindcasting is taken from measured data. In this study, the reforecasting period was kept equal to ~~that~~ the one employed in the studies~~y~~ of ~~(~~López López et al.~~,~~ (2014) and ~~(~~Dogulu et al.~~,~~ (2015). The chosen period is from 1 January 2006 ~~through~~ to 7 March 2013. Data in the period till 6 March 2007 is used for the model spin up. The remaining period is used for the calibration and validation of the uncertainty estimation techniques. Forecasts are made on a 12 hourly basis – at 8:00 and 20:00 daily, up to a lead time of 48 hr. kNN resampling was applied for forecasts at 10 different lead times: 1, 3, 6, 9, 12, 18, 30, 36, 42, and 48 hr. To choose an input variable vector for kNN resampling, correlation analysis was done between residual error and contenders for input variable vector space, namely simulated water level (H^{sim}), measured water level (H^{obs}) and residual error (E^{obs}) from various time steps t. The analysis was done to assist in a manual selection of input variable vectors. The correlation between residual error and water level reduces fast with time lag between the two time series. Therefore it is enough to choose relatively simple and small dimensional input variable vector spaces. For lead times, l, up to 6 hours we chose:

$$\boldsymbol{v} = [H_{\mathrm{t}}^{\mathrm{sim}}, H_{\mathrm{t-l}}^{\mathrm{obs}}, e_{\mathrm{t-l}}^{\mathrm{obs}}] \tag{15}$$

For higher lead times, uncertainty has only been conditioned on H_t^{sim} as the residual error becomes less and less correlated with variable values at measured several hours behind the prediction time.

$$\boldsymbol{v} = [H_{\mathrm{t}}^{\mathrm{sim}}] \tag{16}$$

Ninety nine values of residual errors were sampled from the nearest neighbourhood to generate an empirical distribution at each prediction step. This allowed us to get the 'resolution' of 1 percentile in the generated empirical distribution. To develop confidence in the chosen value of k, we checked for a few prediction steps how sensitive the generated empirical distribution is to the value of k. Four different instances of $\mathbf{v_t}$ were chosen. Each instance represents a prediction step in the input variable vector space (the red circle in Figure 1), with different hydrologic conditions. The plots of the cumulative mean square difference between pdfs of varying k were generated. Cumulative mean square difference (Eq. (17)) ~~just~~ serves as an index to show how much the empirical pdfs change with changing k. We get a decreasing slope with increasing k. It shows that the pdfs become almost identical for values of k around 100. If $\mathrm{P_{k_i}}(\mathrm{e})$ is the probability density for a residual error e calculated through $\mathrm{k_i}$ nearest neighbours using kNN resampling, for probability functions corresponding to discrete bin size $\Delta\mathrm{e}$, the cumulative difference is defined as:

$$\text{Cumulative Difference} = \sum_{\mathrm{k_i=10}}^{\mathrm{k_i=k}} \sum_{\text{first e bin}}^{\text{last e bin}} [\Delta e \cdot P_{\mathrm{k_i+1}}(e) - \Delta\mathrm{e} \cdot P_{\mathrm{k_{i-1}}}(e)]^2 \tag{17}$$

The various values of $\mathrm{k_i}$ that were tested are -10, 30, 50, 70, 90, 100, 110, 130 and 150. Using the information from Figure 3, a value of k = 99 does not seem to be heavily effected by sampling errors. Nevertheless, it is not a mathematically calibrated value of k and therefore is likely to be sub-optimal. However, it should still be able to provide reasonably representative samples from the error distribution, as is suggested by Figure 3.

3.1.3 Results

Figure 4 shows two hydrographs for the same event, where model predictions were made at different lead times. From the graph of lead time 48 hr it is evident that the error quantiles that kNN resampling produces are not forced to have zero mean. Therefore the model prediction can sometimes lie outside the predicted quantiles. This is because kNN resampling learns from past instances where the model has consistently under or overpredicted the flow, so it corrects for this bias. The hydrographs capture the low flows and the peaks well. It can also be seen that for high flows the errors are usually higher than for medium and low flows. The residual error distribution is thus heteroscedastic, i.e. the variance depends on the magnitude of the predicted flow. The autocorrelation can be checked by plotting errors versus time. Whereas performance of an error model with regard to heteroscedasticity can be estimated by plotting reliability diagrams for different magnitudes of flow, which would mean different water levels in this case.

The plotting of error time series (Figure 5)~~.–~~ for various lead times shows some recurring trends across all the three subcatchments. The errors are small for small lead time forecasts and the spread of error time series increases with increasing lead time. Moreover, the errors do not look ~~to be~~ autocorrelated for smaller lead times, whereas for the higher lead times autocorrelation becomes more prominent. This can be ascribed to the memory of the hydrologic system. If the system response is higher than what the model simulates for a particular lead time, then the system response is likely to be higher for the next time step as well. As the errors become larger, they tend to lose their independence property. This is captured by the error samples generated by kNN resampling as well. The rate at which autocorrelation deteriorates for observed residual errors corresponds well to the kNN resampling error samples' autocorrelation. ~~(Figure 5). –~~It can be seen that kNN resampling preserves the autocorrelation in the error time series without using an autoregressive model.

To check the performance of kNN resampling for various flow magnitudes, the simulation values were divided into low and high flows - the lowest and the highest 10 percent of water levels simulated in the validation phase respectively. The reliability diagrams (Figure 6) shows that the overall performance of error quantiles for all water levels is good for low and medium lead times. The reliability decreases with high lead times (24 hr and above). The reliability plots show that kNN resampling performs better for high flows compared to lows flows, even for higher lead times. For lows flows and high lead times, the forecast probability of non-exceedance is higher than the observed relative frequency. Nevertheless, from 0.90 probability of non-exceedance and above, the reliability curve comes back to the desired 45^0 line. For flood forecasting it is important to model the high and medium flows well. kNN resampling delivers quite reliable quantiles for such flow regimes. The deteriorating model performance with higher lead times gets reflected in the performance of kNN resampling quintiles as well.

To assess the performance of kNN resampling relative to other established post-~~-~~processor uncertainty estimation techniques, comparisons with QR and UNEEC have been carried out. The results for QR have been taken from López López et al. (2014) and the results for UNEEC – from Dogulu et al. (2015). QR results for uncertainty estimation were available for the all the lead times as done using kNN resampling~~,~~ and, from UNEEC, ~~—~~only for lead times of 1, 3, 6, 9, 12, and 24 hr. Values of PICP and MPI are shown in Figure 7, together with results from UNEEC and QR. The Alpha Index (α) is reported for several lead times in Table 2. As expected, the ~~mean prediction interval (MPI)~~MPI of all the uncertainty estimation techniques increases with increasing lead time. Comparison between kNN resampling and QR has been made for 3 locations and~~,~~ 10 lead times in the validation period. Model simulations were run two times each day. Verification indices for uncertainty analysis were calculated separately for each lead time and each location. Considering 90% and 50% quantile as two prediction intervals, this allowed for the evaluation of PICP and MPI 60 times (Figure 7). kNN resampling has higher

PICP in 67% of the cases and a smaller MPI for 73% of the cases. A comparison between kNN resampling and UNEEC was made for 3 locations and~~, only~~ 5 lead times for the validation. For each location and each lead time, the 90% and 50% quantiles were generated, which allowed for the evaluation of PICP and MPI 30 times (Figure 7). The PICP of kNN resampling is higher in 60% of the cases and the MPI is smaller in 36% cases. Based on these results we concluded that, for this case study, kNN resampling generally produces narrower confidence bands and provides a better coverage of the probability distribution than the other methods in the majority of forecasts, especially showing better performance for the larger lead times.

3.2 River Brue

3.2.1 Catchment description

River Brue, located in the south west of England, has a history of severe flooding. The test forecast location used in this study is Lovington, where the upstream catchment area is 135 km^2 (Figure 8). The catchment is predominantly rural and the soil consists of clay and sand. This kind of soil and the modest relief give rise to a slowly responsive flow regime. The mean annual rainfall in the catchment is 867 mm, the mean river flow is 1.92 m^3/s. and has a maximum flow of 39.58 m^3/s. This catchment has been extensively used for research on weather radar, quantitative precipitation forecasting and hydrologic modelling.

3.2.2 Experimental setup

For Brue catchment the simplified version of the HBV rainfall-runoff model has been used (Bergström, 1976). HBV-96 model is a lumped conceptual model (Lindström et al., 1997). Like most other conceptual models, HBV consists of subroutines for snow accumulation and melt, soil moisture accounting and surface runoff, and employs a simple routing scheme. The input for the HBV model consists of precipitation (basin average), air temperature and potential evapotranspiration (estimated by modified Penmann method using automatic weather data available). Historical input data is available for a period of 1994-1996. Predictions are only made for 1 hr lead time. Uncertainty analysis is done for a chosen period from the 24th June of 1994 ~~through~~ to 31st of May 1996. Hindcasts were made on a daily basis, using a warm state from a historical run. The hindcasts were split into calibration and validation set at 24th June 1995 for the uncertainty estimation techniques. The calibration data set was used to calibrate (train) UNEEC and QR, ~~whereas~~ and for the ~~kNN~~ resampling of errors using kNN algorithm. ~~was allowed to learn only from this same calibration data set~~The resampled errors were used to estimate prediction intervals for the ~~vectors~~ predictions from the validation data set. Each of the two data sets represents almost a full year of observations. Three input variable vectors were~~as~~ chosen based on the results of correlation analysis, from simple to complex. This allows to study the dependence on the choice of search space. Input variable vector

(ivv) 3 for kNN resampling and UNEEC is same for this comparison, whereas, QR only uses Q_{sim} ~~To make an appropriate comparison, the same variables have been used for kNN resampling as used for UNEEC in~~ (Dogulu et al., 2015). The three input variable vectors ~~is~~used are:

$$\boldsymbol{v}_{(\text{ivv1})} = [\, Q^{sim}_{\text{t}} \,] \qquad (18)$$

$$\boldsymbol{v}_{(\text{ivv2})} = [\, Q^{sim}_{\text{t}}, e^{\text{obs}}_{\text{t}-1} \,] \qquad (19)$$

$$\boldsymbol{v}_{(\text{ivv3})}\ \cancel{\boldsymbol{v}} = [R^{\text{obs}}_{\text{t}-8}, R^{\text{obs}}_{\text{t}-9}, R^{\text{obs}}_{\text{t}-10}, Q^{\text{obs}}_{\text{t}-1}, Q^{\text{obs}}_{\text{t}-2}, Q^{\text{obs}}_{\text{t}-3}, e^{\text{obs}}_{\text{t}-1}, e^{\text{obs}}_{\text{t}-2\cancel{1}}] \qquad (20\cancel{18})$$

where R is the effective rainfall, Q is the discharge and e is the residual error. Considering t as the prediction time, then the subscript of the various input variables represents the time and the superscript *sim* and *obs* means they are simulated and observed values respectively. The number of nearest neighbours was chosen to be 99 and 199, ~~-~~ to analyse its influence on uncertainty quantification~~after a similar manual calibration procedure as for the Upper Severn case study.~~. ~~-~~Uncertainty analysis was done for a calibrated HBV model as well as a model with a unit systematic bias. The bias was introduced to the simulation results of the calibrated model by simple addition. The aim of a biased model for uncertainty quantification using kNN resampling is to assess the performance of kNN resampling when the residuals are not zero mean.

Formatted: Font: Italic

3.2.3 Results

kNN resampling was applied to a single historical simulation and compared to observations. The simulated hydrographs for ~~two events~~highest discharge event with 50% and 90% prediction intervals are shown in Figure 9. The residual distribution of kNN resampling is generally non-zero mean. Therefore we see that the prediction intervals may sometimes deviate from the deterministic model prediction quite significantly. The ability of kNN resampling to search for similar hydrologic conditions, like rainfall, and discharge in the past, and learn from the residuals, allows it to make more representative error distributions. For example, in Figure 9, the falling limb of hydrograph ~~in event 2~~ shows that the prediction band generated by kNN resampling captures the observed flow ~~almost perfectly~~for input variable vector 2 and 3, even though the model shows a noticeable mismatch with the measurements. This can be explained by considering the history of errors that the model made during such hydrologic conditions in the past. And as kNN resampling learns that the model consistently underestimates in such cases, the corresponding error distribution corrects for this bias. The results of the PICP and MPI are shown in T~~t~~able 3~~2~~ together with results from UNEEC and QR (Dogulu et al., 2015). ~~-~~As can be seen from the table, kNN resampling's performances ~~s~~ is comparable ~~better~~to ~~-~~that ~~n~~of ~~the~~ UNEEC and QR for this case study. The prediction intervals generated by kNN resampling are smaller, compared to the other two uncertainty estimation techniques, while the coverage probability is similar. It indicates that kNN resampling is able to learn well from past data and condition the probability of residual errors well. The Alpha Index for the validation phase is also high (0.96). It is also noticed that three different input variable vectors show different degrees of performance (Figure 9). The past errors, e^{obs}_{t-1}, seem to be informative in this case, providing very narrow conditional error probabilities.

Apart from evaluating the usability of kNN resampling for calibrated models, the performance of kNN resampling quantiles generated by kNN resampling for a model with systematic bias was also checked. Figure 10 shows that the performance of kNN resampling does not diminish under systematic bias. The reliability of the generated quantiles remains almost unfazed. As a systematic bias will not affect the autocorrelation structure of the residual errors, the autocorrelation of error samples generated through kNN resampling also remains unchanged. Nevertheless, we see a shift in the mean of the sample time series, which is roughly equal to unity. The reliability of quantiles generated using kNN resampling for high flows (highest 10% in the validation period) is poorer~~as good~~ than~~as~~ for all flows. ~~Thus kNN resampling maintains its accuracy for flow regimes that are of interest in flood forecasting, like it did for Upper Severn catchments.~~ The invariance of kNN resampling performance to model bias ~~and high predictand magnitudes of~~ makes it a robust post-~~-~~processor ~~uncertainty estimation~~ technique, however, unlike in the case of Upper Severn subcatchments, the technique's performance diminishes for high flows. ~~.~~

4 Discussion and conclusions

The application of kNN resampling to two cases studies shows that the forecast uncertainty intervals are relatively narrow and still capture the observations well. The expected increase of uncertainty for longer lead times is also reproduced well and the probability coverage of kNN resampling remains good as verified from historical observations. This is in accordance with previous research (Sikorska et al., 2015). The error samples generated by kNN resampling reproduce two important characteristics of residual errors in hydrologic models namely autocorrelation and heteroscedasticity. Also, for applications to flood modelling, the high flows are most important and the uncertainty quantification by kNN resampling for ~~both case studies,~~ Upper Severn ~~and Brue,~~ shows reasonable reliability for this high flow regime. For Brue, the performance is poorer . This can be attributed to the inadequacy of representative high flows in the calibration phase in combination with the choice of the input variable vector. The highest flow is calibration time series is 15.4 m^3/s whereas in validation time series it is 29.9 m^3/s. ~~Moreover, i~~It is also shown that the technique is generally robust to the performance of the underlying deterministic model. If the model has systematic biases, kNN resampling learns from the past errors of the model and recreates the systematic bias in the empirical error distributions mean, thus maintaining the performance of prediction intervals. Our results on systematic error correction by kNN resampling substantiate the findings from previous research on forecast updating using kNN (Akbari and Afshar, 2014). These findings from this study are confirmed by ~~two~~ three quantitative indicators of forecast reliability. The comparison of kNN resampling uncertainty estimates to those generated by QR and UNEEC show that the mean prediction intervals (MPI) generated by kNN resampling are generally smaller. Significantly smaller MPI using kNN resampling, as in the case of Brue, is in part, due to the conditioning on input variable vector, as compared to UNEEC and QR. As the value of k in this study has been restricted to 99 and 199, the error distribution tend to be much narrower than the marginal error distribution. The conditional distribution will turn into marginal distribution when the number of k is equal to the time steps in the calibration time series. A more quantitative dependence on k value and MPI will need further research. Apart from a narrow MPI, we also find that kNN resampling is generally able to capture the expected ratio of observations within its intervals (PICP) most of the times, or at least be close to the expected value.

As in the case of all other data-driven methods, the applicability of kNN resampling depends on the availability of sufficiently long and representative historical forecasts and observations. The historical series should include several occurrences of forecasting situations that are similar to the current situation. In extreme cases, the kind of kNN search proposed here will select the most similar historical situations which may or may not be representative of the current situation. In contrast to the methods like QR and UNEEC that build explicit predictive regression models which are able to extrapolate for the data which is beyond the limits of the calibration (training set), kNN resampling does not extrapolate. This could be seen as a disadvantage. On the other hand, however, the extrapolation that is done by regression techniques could be also seen as doubtful. It is not a given that the most extreme historical situations are less representative for the uncertainty of an extremely high flow than an extrapolated result. The results ~~of both case studies~~ in this paper show that kNN resampling has a good or poor reliability for the highest values in the validation set, depending on the case study and the choice of input variable vector. Due to the non-parametric nature of kNN resampling, the increasing variance of residual errors for higher values of predictand is generally adequately taken into account.

As kNN resampling, like other post ~~-~~processors, learns about the residual error process from the past, the historical records should be representative of the current forecast conditions. In changing conditions, this may not be true. Changing conditions may be caused, for example due to climate change or more local changes in the catchment like deforestation, dam building etc. This is a common problem for all data-driven statistical estimators and not unique to kNN resampling. Care needs to be taken to use data time series which do not outrightly violate the assumptions regarding the invariance of catchment and climate.

One of the few calibration parameters of kNN resampling is the number of nearest neighbours k. In this study, k has been chosen by a simple heuristic technique. For optimal performance, it would be advisable to calibrate k for each application in a more systematic way. We do show for Brue that the sensitivity of the uncertainty intervals to the value of k is not significant, when changing it from 99 to 199. However, ~~W~~we also expect that the optimal value of k will depend on the length of the historical data series and on the uncertainty quantiles of interest. In the context of search space,~~Also, i~~ in this research, the input variable vector has been chosen by correlation analysis. It can be recommended to use more sophisticated procedures for real life applications, which can capture the non-linear dependence between the error process and input variable vector candidates. ~~It is foreseen that i~~Improvements in performance can possibly be achieved by seeking a better set of input variables for each forecast location and lead time of interest.

In conclusion, kNN resampling can be considered as a relatively simple machine learning technique to predict hydrologic residual uncertainty. The errors from the similar hydrologic conditions in the past are used as samples for the residual error probability distribution and the samples are collected by a k nearest neighbour search. The application of this technique to case studies Brue and Upper Severn subcatchments has shown promising results. In comparison to many other data driven techniques, kNN resampling has the advantage of avoiding assumptions about the nature of the residual error distribution: the instance-based learning approach is non-parametric and non-regressive and requires little calibration. The method was shown to be able to quantify hydrologic uncertainty to an accuracy that is comparable to other techniques like QR and UNEEC. Given the relatively small effort in setting up the method, the performance of kNN resampling in uncertainty quantification is more than acceptable when compared to other post-~~-~~processor error models.

User interface

A website has been developed as part of this research to help generate uncertainty ~~estimation~~ intervals using kNN resampling for a given time series of predictions. Address: www.modeluncertainty.com

Acknowledgements

The authors would like to acknowledge Bonneville Power Administration, Portland, USA, for supporting this research. The UK Environment Agency is acknowledged for provision of the data for the case studies described in this manuscript. Many thanks to Nilay Dogulu, Patricia López López, and Marijn Swenne for their help during the course of this research. We are thankful to Sven Eggimann and H. Badger for helping structure the paper. Part of this study was supported by the EC FP7 project WeSenseIt (Citizen Observatory of Water), grant agreement no 308429 and project QUICS (Quantifying Uncertainty in Integrated Catchment Studies), grant agreement no. 607000.

References

Akbari, M., and Afshar, A.: Similarity-based error prediction approach for real-time inflow forecasting, Hydrol Res, 45, 589, 10.2166/nh.2013.098, 2014.

Arnal, L., Ramos, M.-H., Coughlan de Perez, E., Cloke, H. L., Stephens, E., Wetterhall, F., van Andel, S. J., and Pappenberger, F.: Willingness-to-pay for a probabilistic flood forecast: a risk-based decision-making game, Hydrol Earth Syst Sc, 20, 3109-3128, 10.5194/hess-20-3109-2016, 2016.

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

Beckers, J. V. L., Weerts, A. H., Tijdeman, E., and Welles, E.: ENSO-conditioned weather resampling method for seasonal ensemble streamflow prediction, Hydrol Earth Syst Sc, 20, 3277-3287, 10.5194/hess-20-3277-2016, 2016.

Benke, K. K., Lowell, K. E., and Hamilton, A. J.: Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model, Mathematical and Computer Modelling, 47, 1134-1149, 10.1016/j.mcm.2007.05.017, 2008.

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

Beven, K., and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279-298, 10.1002/hyp.3360060305, 1992.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, Journal of Hydrology, 298, 242-266, 10.1016/j.jhydrol.2004.03.042, 2004.

Coccia, G., and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrol Earth Syst Sc, 15, 3253-3274, 10.5194/hess-15-3253-2011, 2011.

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, Hydrol Earth Syst Sc, 17, 4209-4225, 10.5194/hess-17-4209-2013, 2013.

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L., and Tait, S.: Assessing uncertainties in urban drainage models, Physics and Chemistry of the Earth, Parts A/B/C, 42-44, 3-10, 10.1016/j.pce.2011.04.007, 2012.

Dogulu, N., Lopez, P. L., Solomatine, D. P., Weerts, A. H., and Shrestha, D. L.: Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments, Hydrol Earth Syst Sc, 19, 3181-3201, 10.5194/hess-19-3181-2015, 2015.

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W., and Deletic, A.: Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling, Water Research, 46, 2545-2558, 10.1016/j.watres.2012.02.009, 2012.

EA: Environment Agency: River levels: Midlands, available at: http://www.environment-agency.gov.uk/homeandleisure/floods/ riverlevels/, (last access: 1 October 2013), 2009.

Evin, G., Kavetski, D., Thyer, M., and Kuczera, G.: Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, Water Resources Research, 49, 4518-4524, 10.1002/wrcr.20284, 2013.

Fernando, T. M. K. G., Maier, H. R., and Dandy, G. C.: Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach, Journal of Hydrology, 367, 165-176, 10.1016/j.jhydrol.2008.10.019, 2009.

Freer, J., Beven, K., and Ambroise, B.: Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach, Water Resources Research, 32, 2161-2173, 10.1029/95WR03723, 1996.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resources Research, 34, 751-763, 10.1029/97wr03495, 1998.

Hoss, F., and Fischbeck, P. S.: Performance and robustness of probabilistic river forecasts computed with quantile regression based on multiple independent variables, Hydrology and Earth System Sciences, 19, 3969-3990, 10.5194/hess-19-3969-2015, 2015.

Jules, J. B., and Buishand, T. A.: Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation, Climate Research, 25, 121-133, 2003.

Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resources Research, 35, 2739-2750, 10.1029/1999WR900099, 1999.

Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, Journal of Hydrology, 249, 2-9, http://dx.doi.org/10.1016/S0022-1694(01)00420-6, 2001.

Laio, F., and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11, 1267-1277, 10.5194/hess-11-1267-2007, 2007.

Lall, U., and Sharma, A.: A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series, Water Resources Research, 32, 679-693, 10.1029/95wr02966, 1996.

Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, Journal of Geophysical Research, 114, 10.1029/2008jd010969, 2009.

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.

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, Hydrology and Earth System Sciences, 18, 3411-3428, 10.5194/hess-18-3411-2014, 2014.

Lopez, P. L., 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 Sc, 18, 3411-3428, 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.

Montanari, A., and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resources Research, 40, n/a-n/a, 10.1029/2003wr002540, 2004.

Pianosi, F., and Raso, L.: Dynamic modeling of predictive uncertainty by regression on absolute errors, Water Resources Research, 48, n/a-n/a, 10.1029/2011wr010603, 2012.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Monthly Weather Review, 133, 1155-1174, 10.1175/mwr2906.1, 2005.

Rajagopalan, B., and Lall, U.: A_k-nearest-neighbor simulator for daily precipitation and other weather variables, Water Resources Research, 35, 3089-3101, 10.1029/1999wr900028, 1999.

Refsgaard, J. C., van der Sluijs, J. P., Hojberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process - A framework and guidance, Environ Modell Softw, 22, 1543-1556, 10.1016/j.envost.2007.02.004, 2007.

Reggiani, P., Renner, M., Weerts, A. H., and van Gelder, P. A. H. J. M.: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, Water Resources Research, 45, n/a-n/a, 10.1029/2007wr006758, 2009.

Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Spörri, C., Tockner, K., and Truffer, B.: Concepts of decision support for river rehabilitation, Environ Modell Softw, 22, 188-201, 10.1016/j.envsoft.2005.07.017, 2007.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, n/a-n/a, 10.1029/2009WR008328, 2010.

Roscoe, K. L., Weerts, A. H., and Schroevers, M.: Estimation of the uncertainty in water level forecasts at ungauged river locations using quantile regression, International Journal of River Basin Management, 10, 383-394, 10.1080/15715124.2012.740483, 2012.

Schoups, G., and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, Water Resources Research, 46, n/a-n/a, 10.1029/2009wr008933, 2010.

Shrestha, D. L., and Solomatine, D. P.: Data-driven approaches for estimating uncertainty in rainfall-runoff modelling, International Journal of River Basin Management, 6, 109-122, 10.1080/15715124.2008.9635341, 2008.

Sikorska, A. E., Alberto, M., and Demetris, K.: Estimating the Uncertainty of Hydrological Predictions through Data-Driven Resampling Techniques, 10.1061/(ASCE)HE.1943-5584.0000926, 2015.

Solomatine, D. P., Maskey, M., and Shrestha, D. L.: Instance-based learning compared to other data-driven methods in hydrological forecasting, Hydrological Processes, 22, 275-287, 10.1002/hyp.6592, 2008.

Solomatine, D. P., and Ostfeld, A.: Data-driven modelling: some past experiences and new approaches, Journal of Hydroinformatics, 10, 3-22, 10.2166/hydro.2008.015, 2008.

Solomatine, D. P., and Shrestha, D. L.: A novel method to estimate model uncertainty using machine learning techniques, Water Resources Research, 45, n/a-n/a, 10.1029/2008wr006839, 2009.

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, International Journal of River Basin Management, 6, 123-137, 10.1080/15715124.2008.9635342, 2008.

van Andel, S. J., Weerts, A., Schaake, J., and Bogner, K.: Post-processing hydrological ensemble predictions intercomparison experiment, Hydrological Processes, 27, 158-161, 10.1002/hyp.9595, 2013.

van der Vaart, a W.: Asymptotic Statistics, Asymptot. Stat., 3, 443, doi:10.2307/2530729, 1998.

Verkade, J. S., Brown, J. D., Reggiani, P., and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, Journal of Hydrology, 501, 73-91, 10.1016/j.jhydrol.2013.07.039, 2013.

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/ (last access: 1 October 2013), HR Wallingford, Hydraluic Unit, Wallingford, UK, 1997.

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 Sc, 15, 255-265, 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, Environmental Modelling & Software, 40, 65-77, 10.1016/j.envsoft.2012.07.010, 2013.

Table 1: Basin information for Upper Severn subcatchments (EA, 2013 and Marsh, T. J. and Hannaford, J., 2008)

Catchment	Area (km^2)	Mean Annual Rain (mm)	Mean Flow (m^3/s)	Max Water level (m)
Llanerfyl	125	1077	>10	3.59
Llanyblodwel	229	1267	6.58	2.68
Yeaton	180.8	767	1.6	1.13

Table 2: Index Alpha (α) for different lead times of Upper Severn subcatchments

Lead Time (h)	1	12	24	48
Llanerfyl	0.92	0.87	0.79	0.64
Llanyblodwel	0.93	0.95	0.93	0.90
Yeaton	0.97	0.94	0.94	0.75

Formatted Table

Table 3~~2~~: Performance of various uncertainty estimation techniques for Brue catchment. For kNN resampling and UNEEC the same input variable vector is used (Eq. (20)). For QR only Q_{sim} is used.

PICP (Expected 90%)			MPI (m^3/s)		
UNEEC	QR	kNN	UNEEC	QR	kNN

Calibration	91.19	90.00	~~95.11~~86.3	1.58	1.69	0.~~49~~51
Validation	88.29	82.33	~~92.15~~84.42	1.37	1.39	0.~~39~~21

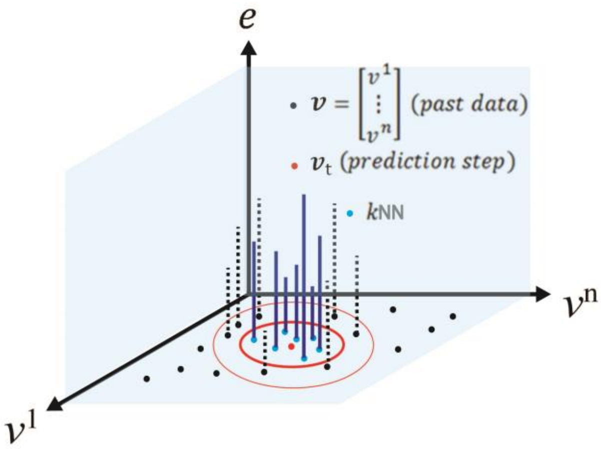

e
$v = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}$ (past data)
v_t (prediction step)
kNN
v^n
v^1

Figure 1. Dependence of error samples on the value of k. For larger values of k, points are at a greater distance from v_t (the prediction step), thus compromising the conditioning of the residual error probability distribution on v_t (Eq. (5)).

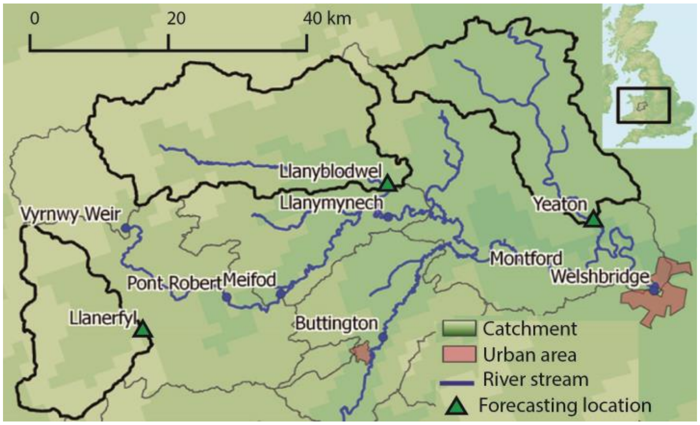


Figure 2. Upper Severn subcatchments with gauging stations (From López López et al., 2014)

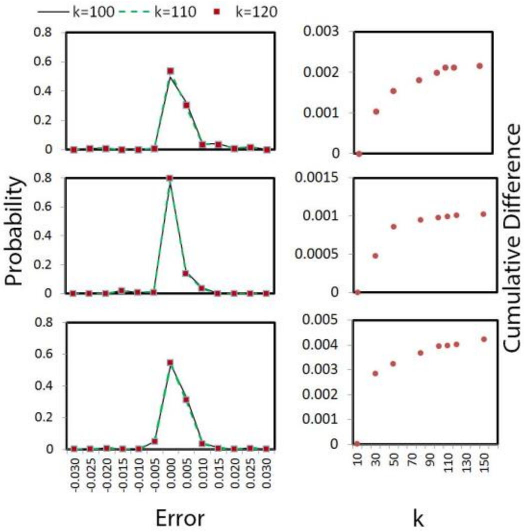


Figure 3. **Dependence of residual error probability function on the value of k for three didactic values of v_t (each row). The probability is computed for error bins of size 0.005 units each. The graphs show that for k from around 90 to 120, the corresponding empirical error distributions become almost identical.**

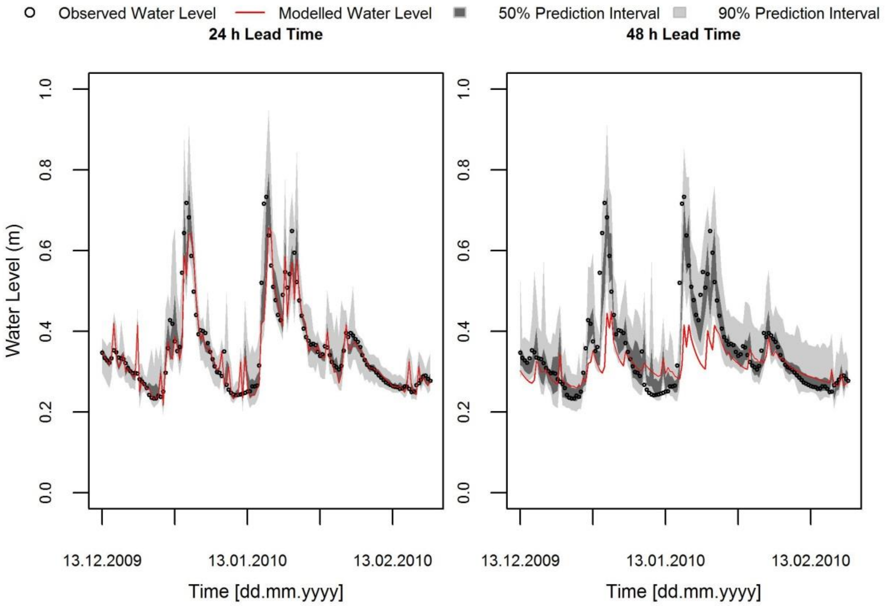


Figure 4. Prediction intervals for Yeaton catchment using kNN resampling. The hydrographs are shown for the two different lead times. 50% prediction interval is the interval between 25% and 75% quantiles of residual error, and 90% quantile is the interval between 5% and 95% quantiles. The reporting time interval is 12 hours.

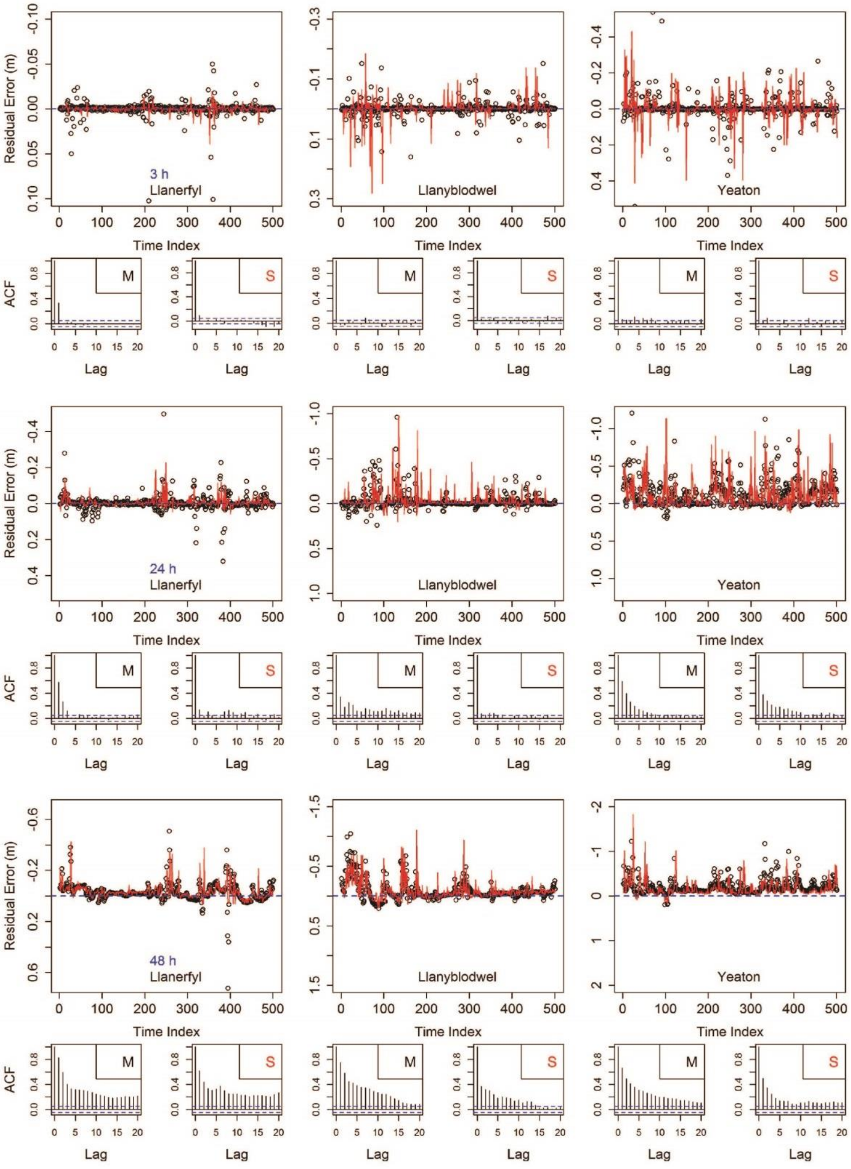
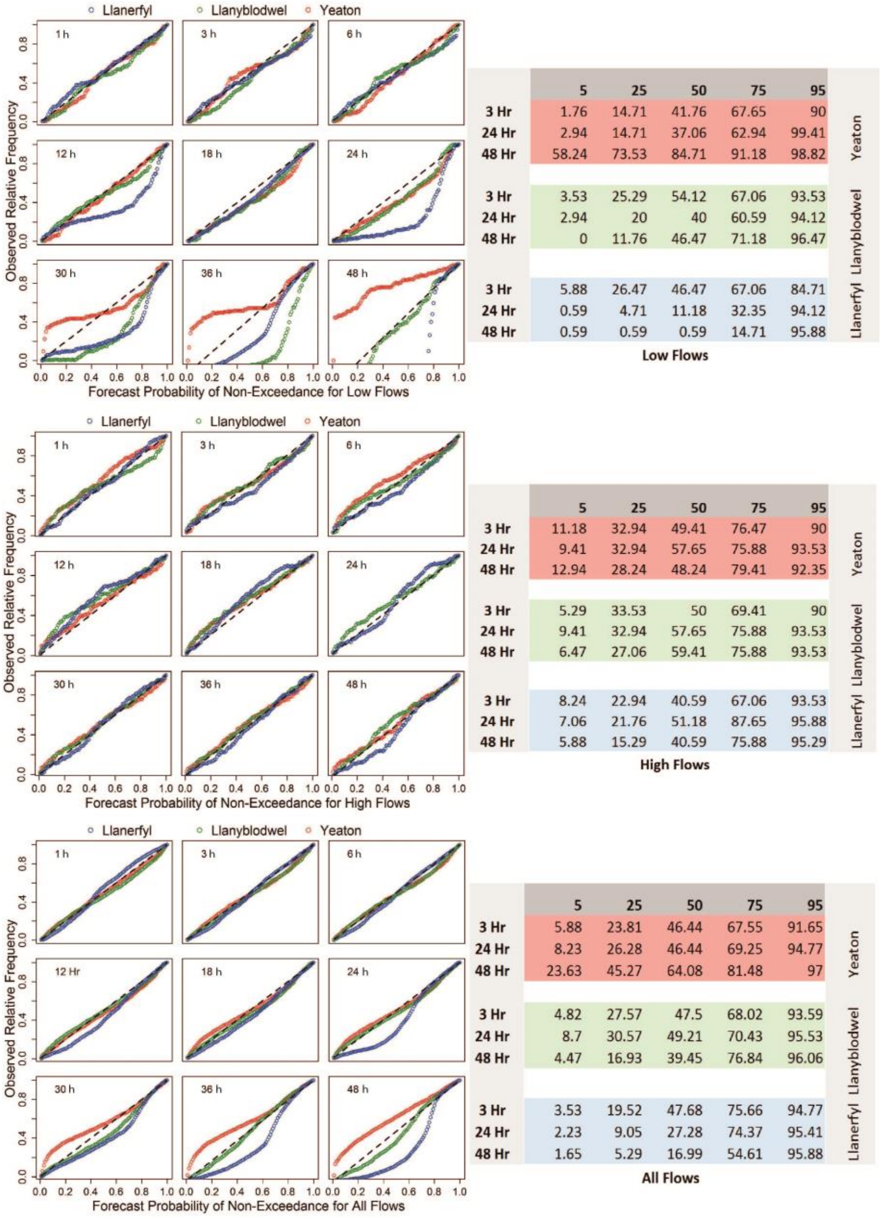


Figure 5. **Plots of error samples and their autocorrelation (ACF). The error time series generated using kNN resampling are in red. Black circles represent the observed errors, i.e. obtained after measuring water level and comparing it to simulated water level. M stands for measured and S for simulated. The lead time for each row of plots is given in blue.**



	5	25	50	75	95	
3 Hr	1.76	14.71	41.76	67.65	90	Yeaton
24 Hr	2.94	14.71	37.06	62.94	99.41	
48 Hr	58.24	73.53	84.71	91.18	98.82	
3 Hr	3.53	25.29	54.12	67.06	93.53	Llanyblodwel
24 Hr	2.94	20	40	60.59	94.12	
48 Hr	0	11.76	46.47	71.18	96.47	
3 Hr	5.88	26.47	46.47	67.06	84.71	Llanerfyl
24 Hr	0.59	4.71	11.18	32.35	94.12	
48 Hr	0.59	0.59	0.59	14.71	95.88	

Low Flows

	5	25	50	75	95	
3 Hr	11.18	32.94	49.41	76.47	90	Yeaton
24 Hr	9.41	32.94	57.65	75.88	93.53	
48 Hr	12.94	28.24	48.24	79.41	92.35	
3 Hr	5.29	33.53	50	69.41	90	Llanyblodwel
24 Hr	9.41	32.94	57.65	75.88	93.53	
48 Hr	6.47	27.06	59.41	75.88	93.53	
3 Hr	8.24	22.94	40.59	67.06	93.53	Llanerfyl
24 Hr	7.06	21.76	51.18	87.65	95.88	
48 Hr	5.88	15.29	40.59	75.88	95.29	

High Flows

	5	25	50	75	95	
3 Hr	5.88	23.81	46.44	67.55	91.65	Yeaton
24 Hr	8.23	26.28	46.44	69.25	94.77	
48 Hr	23.63	45.27	64.08	81.48	97	
3 Hr	4.82	27.57	47.5	68.02	93.59	Llanyblodwel
24 Hr	8.7	30.57	49.21	70.43	95.53	
48 Hr	4.47	16.93	39.45	76.84	96.06	
3 Hr	3.53	19.52	47.68	75.66	94.77	Llanerfyl
24 Hr	2.23	9.05	27.28	74.37	95.41	
48 Hr	1.65	5.29	16.99	54.61	95.88	

All Flows

Figure 6. Reliability diagram from Upper Severn subcatchments for high, low and all flows. (Llanerfyl – blue, Llanyblodwel – green, Yeaton - red).

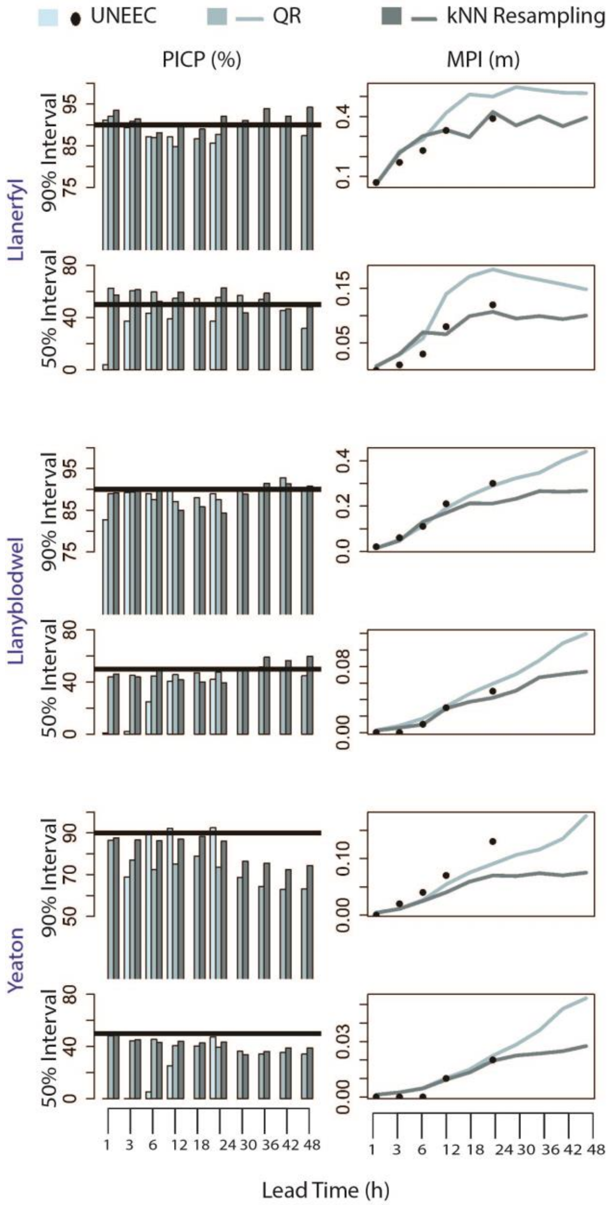


Figure 7. PICP and MPI comparison for Upper Severn subcatchments.

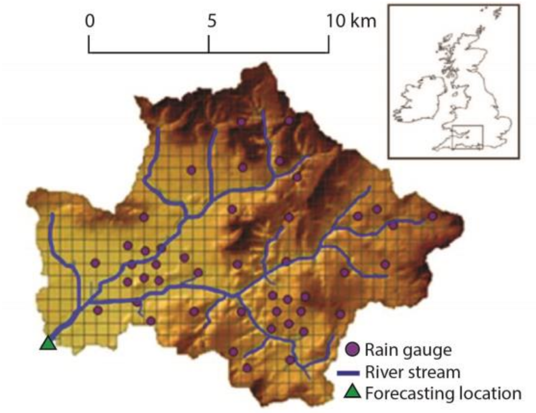
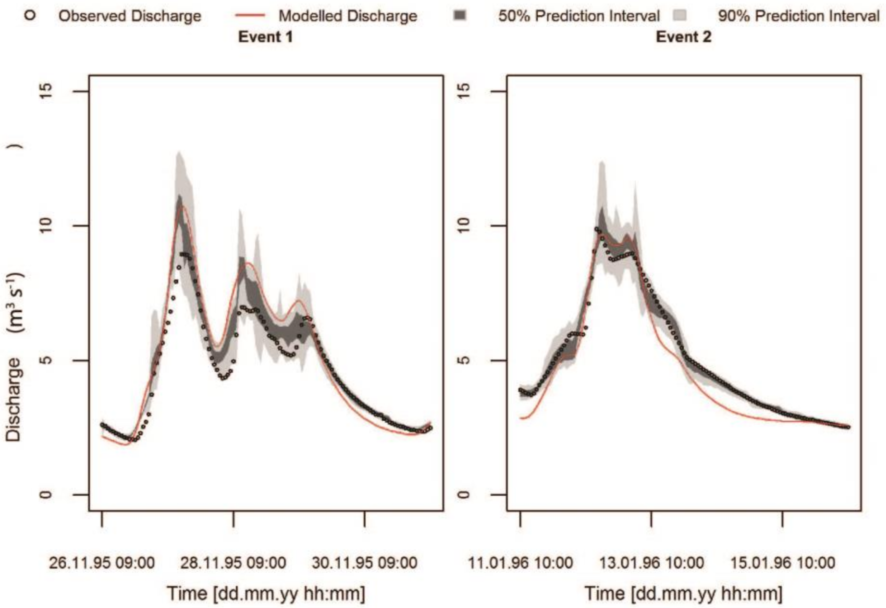
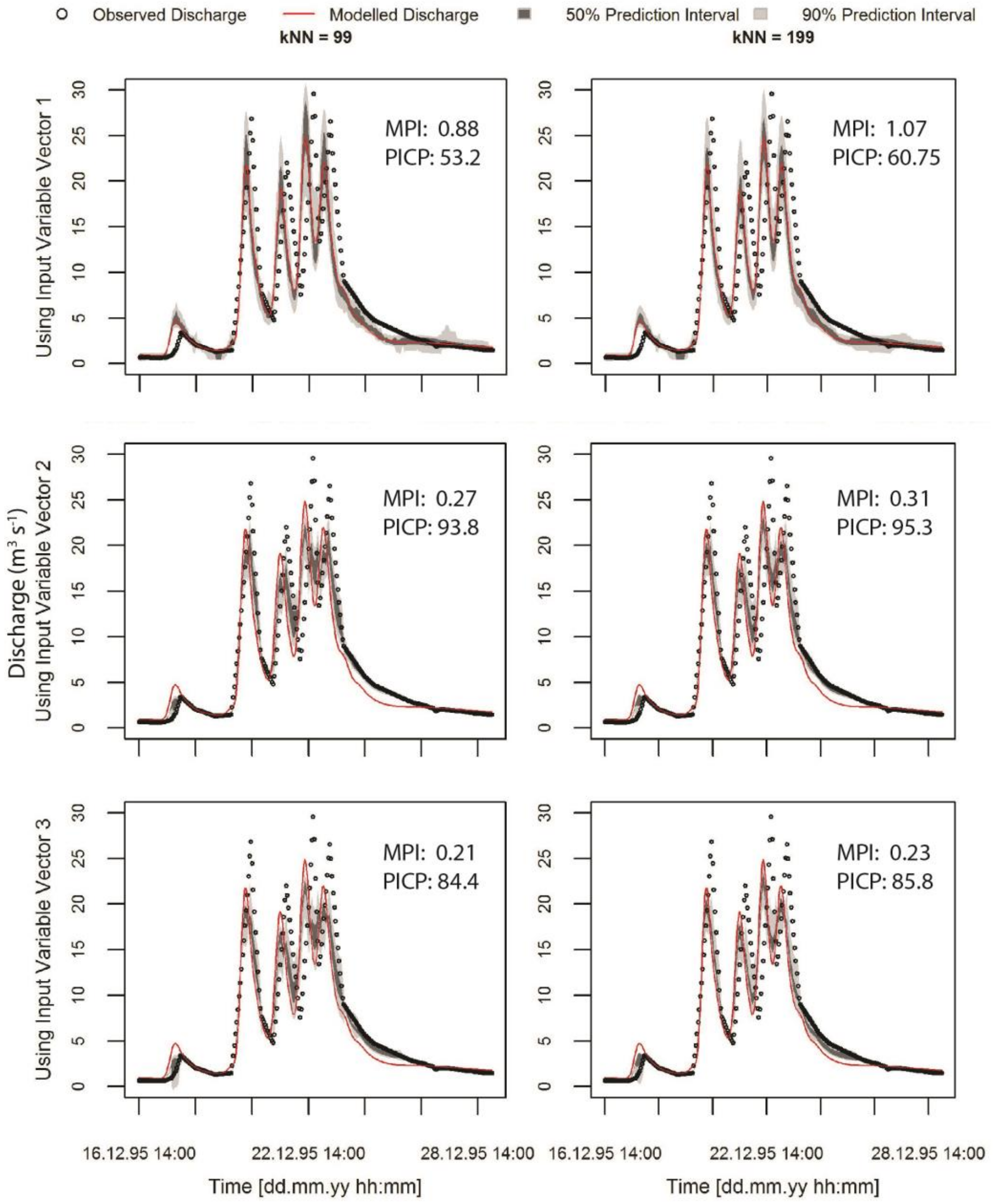


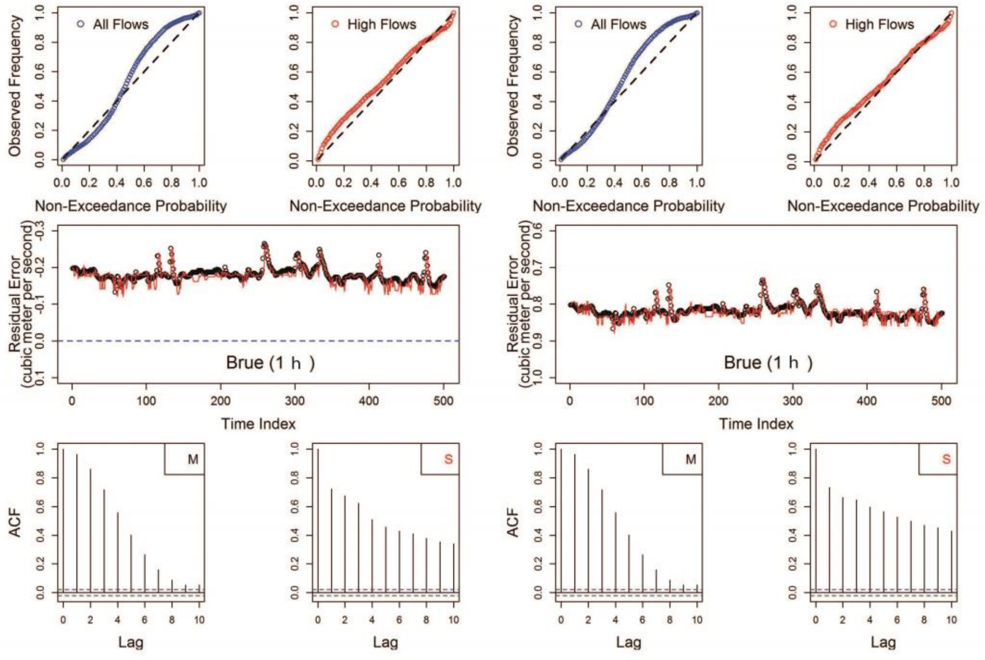
Figure 8. Brue catchment (from Shrestha and Solomatine, 2008)


○ Observed Discharge
Modelled Discharge
50% Prediction Interval
90% Prediction Interval
Event 1
Event 2
Discharge (m³ s⁻¹)
0
5
10
15
26.11.95 09:00
28.11.95 09:00
30.11.95 09:00
11.01.96 10:00
13.01.96 10:00
15.01.96 10:00
Time [dd.mm.yy hh:mm]
Time [dd.mm.yy hh:mm]



○ Observed Discharge
Modelled Discharge
▨ 50% Prediction Interval
▨ 90% Prediction Interval
kNN = 99
kNN = 199
Using Input Variable Vector 1
Discharge (m^3 s^{-1})
Using Input Variable Vector 2
Using Input Variable Vector 3
0
5
10
15
20
25
30
MPI: 0.88
PICP: 53.2
MPI: 1.07
PICP: 60.75
MPI: 0.27
PICP: 93.8
MPI: 0.31
PICP: 95.3
MPI: 0.21
PICP: 84.4
MPI: 0.23
PICP: 85.8
16.12.95 14:00
22.12.95 14:00
28.12.95 14:00
Time [dd.mm.yy hh:mm]

Figure 9. 50% and 90% prediction intervals for Brue catchment using kNN resampling. The hydrographs are shown for two different ~~events~~k values (99, 199) and three different input variable vectors (Eq. (18), Eq. (19) and Eq. (20) for Input Variable Vector 1, 2 and 3 respectively). This is the largest event in the validation time series. (50% prediction interval is the interval between 25% and 75% quantiles of residual error, and 90% quantile is the interval between 5% and 95% quantiles. MPI and PICP correspond to whole validation time series.)



No additional systematic bias

With additional systematic bias

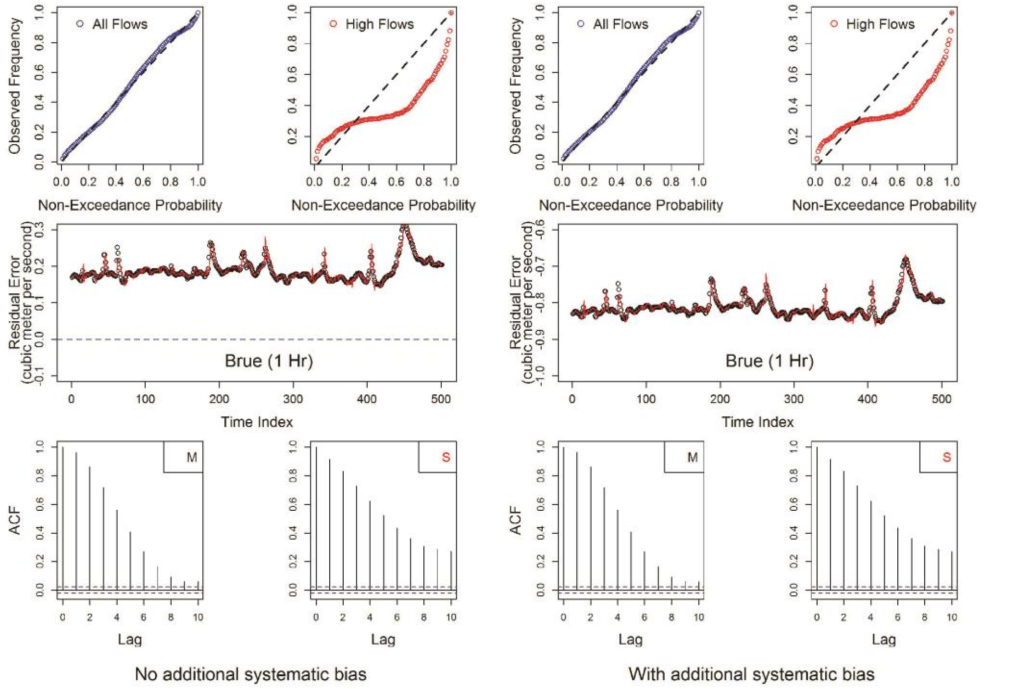


Figure 10. Effect on reliability of quantiles and autocorrelation of error samples on adding a systematic bias to the model artificially. kNN samples, generated using input variable vector 3, are plotted in red and observed errors in black circles. M stands for measured and S for simulated.