Hydrol. Earth Syst. Sci., 15, 255–265, 2011
www.hydrol-earth-syst-sci.net/15/255/2011/
doi:10.5194/hess-15-255-2011

Hydrology and Earth System Sciences

Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales)

A. H. Weerts[1], **H. C. Winsemius**[1], **and J. S. Verkade**[1,2]

[1]Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands
[2]Delft University of Technology, Department of Flood Risk Management and Hydraulic Structures, Delft, The Netherlands

Received: 18 July 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 11 August 2010
Revised: 4 January 2011 – Accepted: 5 January 2011 – Published: 21 January 2011

Abstract. In this paper, a technique is presented for assessing the predictive uncertainty of rainfall-runoff and hydraulic forecasts. The technique conditions forecast uncertainty on the forecasted value itself, based on retrospective Quantile Regression of hindcasted water level forecasts and forecast errors. To test the robustness of the method, a number of retrospective forecasts for different catchments across England and Wales having different size and hydrological characteristics have been used to derive in a probabilistic sense the relation between simulated values of water levels and matching errors. From this study, we can conclude that using Quantile Regression for estimating forecast errors conditional on the forecasted water levels provides a relatively simple, efficient and robust means for estimation of predictive uncertainty.

1 Introduction

Real-time flood forecasting, warning and response systems (often referred to simply as "flood warning systems") aim to give property owners, floodplain residents and responsible authorities time to respond to a flood threat before a critical threshold is exceeded, thus allowing for mitigation of adverse consequences. As such, they constitute a non-structural flood risk management measure. Extending the forecasting lead time allows for time allocation for mitigating actions. A reliable assessment of certainty of predicted events in a real-time context safeguards operational users from issuing false alarms and institutional decision-makers from calling for unwarranted action. Real-time flood forecasting systems are currently operational in many parts of the world, including England and Wales where the National Flood Forecasting System (NFFS) is used by the Environment Agency (Werner et al., 2009).

Following the 2007 summer floods in England and Wales, the Pitt Review (Pitt, 2008) recommended that "The Met Office and the Environment Agency should issue warnings against a lower threshold of probability to increase preparation lead times for emergency responders". This implies that the Environment Agency shifts its flood forecasting operations from a deterministic to a probabilistic approach.

While flood warning systems have the potential to significantly reduce flood risk, uncertainty in water level forecasts may cause imperfect flood warning system performance. This uncertainty has multiple causes including uncertain future boundary conditions of precipitation, evaporation and temperature from numerical weather prediction models, imperfect model schematisations, incorrect parameter estimates and unknown initial states or imperfect estimates thereof. A large body of research over the past decades has attempted to reduce these individual sources of uncertainty. This research includes reduction, characterisation, assessment and modeling of parametric uncertainty, deterministic state updating or data assimilation (Moore, 2007; Serban and Askew, 1991), error correction (Broersen and Weerts, 2005; Shamseldin and O'Connor, 2001), ensemble data assimilation (Clark et al., 2008; Seo et al., 2009; Weerts et al., 2010b), and various probabilistic post-processing techniques

Correspondence to: A. H. Weerts
(albrecht.weerts@deltares.nl)

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

for deterministic forecasts (Krzysztofowicz and Maranzano, 2004; Reggiani and Weerts, 2008; Seo et al., 2006) and for multimodel and/or NWP ensemble prediction based forecasts (Raferty et al., 2003; Reggiani et al., 2009; Sloughter et al., 2007; Todini, 2008; Wood and Schaake, 2008).

While this research has led to considerable progress in reducing uncertainties related to these sources, there will always be a residual uncertainty that cannot be fully eliminated. Although this uncertainty in flood forecasting is widely recognised, many if not most of the existing flood warning systems are based on deterministic forecasts, implying a certain, accurate prediction of water levels. In contrast, probability forecasts explicitly estimate predictive uncertainty about future flows or water levels.

For a number of reasons, the move from deterministic forecasting to probability forecasting constitutes an improvement to flood warning systems. First of all, hydrological forecasts are inherently uncertain. Deterministic forecasting does not acknowledge inherent uncertainties either in a qualitative or quantitative manner, whereas probabilistic forecasts explicitly show the certitude of a prediction. Communicating that level of certitude then allows for the decision to be made by a decision maker rather than a decision being implicitly taken by forecasters. Probability forecasts can then be used to take a risk-based decision, where the consequences of possible outcomes can be weighted by their probability of occurrence function (Raiffa and Schlaifer, 2000; Todini, 2007). Also, depending on these consequences, decision makers can set a threshold of probability against which to decide, thus choosing an appropriate balance between false alarms and missed floods.

Montanari and Brath (2004) report three approaches for estimating predictive uncertainty. The first option is that the model used for forecasting may be structured as a probability model that generates probability distributions. A second option is to estimate predictive uncertainty by analysing the statistical properties of the forecast error series (that is, the difference between the prediction and the observation). The third option is to use simulation and re-sampling techniques, thus applying Monte Carlo analyses. When choosing an approach to be implemented in an operational, real-time flood forecasting system, computational efficiency and data availability need to be taken into account.

The present paper proposes the use of "Quantile Regression" (Koenker, 2005; Koenker and Basset, 1978) as a method to estimate predictive uncertainty. Quantile Regression as applied in this study is an example of the second option mentioned above. The here developed Quantile Regression approach aims to assess the relationship between the hydrological forecast and the associated forecast error. In contrast with "classical" regression methods, Quantile Regression does not optimise on the mean of the dependent variable (the forecast error) but rather on the quantiles (e.g. the median). By thus estimating quantiles, an estimate of the full probability distribution of the forecast error may be achieved. This probability distribution may serve as an estimate for predictive hydrological uncertainty. Other geophysical applications of Quantile Regression can be found in meteorology (Bremnes, 2004), wind forecasting (Bremnes, 2006; Juban et al., 2007; Nielsen et al., 2006) and the prediction of ozone concentrations (Baur et al., 2004; Sousa et al., 2009).

The applicability of Quantile Regression is demonstrated by applying it on a number of catchments in England and Wales. The catchments vary in size and in dominant hydrological processes. A stand-alone version of the National Flood Forecasting System (NFFS, Werner et al., 2004, 2009) was adapted to serve as a prototype of the probabilistic forecasting system.

This paper first describes the theory of Quantile Regression and its application to flood forecasting. Subsequently, the application to the NFFS catchments is described. This is followed by the results and discussion section showing verification metrics. The paper ends with conclusions.

2 Material and methods

2.1 Uncertainty estimation of water level forecasts using quantile regression

With increasing lead time, many sources of uncertainty impact the accuracy of forecasts, with different uncertainty components dominating at different lead times. In an operational setting, forward modelling of all these uncertainties can be infeasible because it requires many data (e.g. meteorological ensemble forecasts) or many model runs (e.g. Beven, 2006).

The stochastic approach used in this study estimates effective uncertainty due to all uncertainty sources. The approach is implemented as a post-processor on a deterministic forecast. We estimate the probability distribution of the forecast error at different lead times, by conditioning the forecast error on the predicted value itself. Once this distribution is known, it can be efficiently imposed on forecast values as a post-processor. We estimate the relationship between the probability distribution of the errors and the forecasted values at a given lead time by means of Quantile Regression. Quantile Regression is a method for estimating conditional quantiles (Koenker, 2005; Koenker and Basset, 1978; Koenker and Hallock, 2001). This requires conditioning of the Quantile Regression relationships on a calibration dataset of forecast values and associated errors at the lead time of interest. To keep the methodology as objective and parsimonious as possible, the degrees of freedom of the Quantile Regression relationships are kept to a minimum by pursuing a linear regression for each quantile of interest. In order to do justify a linear relationship, the heteroscedasticity of the error process, typically associated with rainfall-runoff or hydraulic models, needs to be taken into account. This is done by making the training population of forecast values and associated

errors Gaussian, prior to linear Quantile Regression. In the Gaussian domain, we may assume that the forecast – error relationship may be estimated as a linear function. For each lead time of interest, a different set of Quantile Regression relationships needs to be derived. This is because the magnitude of the effective forecast error increases with lead time. Below, the method to derive the Quantile Regressions is formulated in detail. The method can be used when a sufficiently large sample of hindcasted water levels or flows and concurrent observed values is available.

2.1.1 Transformation of forecast errors from the original domain to the Gaussian domain

Let us denote the process of flow or water levels with a certain lead time as follows:

$$s(t) = \hat{s}(t) + e(t) \tag{1}$$

where s is the real process of river flow or water levels at time t, $\hat{s}$ is the forecasted value and e is the error estimate, all at a certain time t and with a certain lead time (not explicitly denoted here). In practice, we approximate the process s by collecting a population of observations and s by simulation or hindcasting with the lead time of interest, at concurrent time stamps. These estimates can be provided by (a combination of) hydrological and hydraulic models. Let us assume that the error may be estimated by means of a probabilistic error model based on the following functional form:

$$e(t) = f\left[\hat{s}(t)\right] \tag{2}$$

Once this relation is found, it allows a user to apply the error model in any case, without any additional data requirements besides the estimate of the process $\hat{s}$. This is convenient in a real-time operational context, where the availability of additional data besides $\hat{s}$ to estimate uncertainty is not trivial. Now let us assume that the error structure is both ergodic and stationary (i.e. no significant changes in the hydrological or hydraulic processes or measurements have taken place). There is no warrant however, that the error structure is homoscedastic. In fact, residual time series of rainfall-runoff models are known to be heteroscedastic and non-linear in nature. In order derive an objective and reliable probabilistic relation as given in Eq. (2), a transformation of the regressor and the concurrent associated errors to the Gaussian domain is applied through the normal quantile transform (NQT). The NQT (van der Waerden, 1952, 1953a, b) is a non-parametric method to map a variable, having any distribution, to a Gaussian distribution and has been described for hydrological applications by Kelly and Krzysztofowicz (1997). Applications of the NQT in error estimation have been performed by Krzysztofowicz and Maranzano (2004) and by Montanari and Brath (2004). In effect, the plotting positions of the cumulative distribution function of the available samples are associated with their counterparts in the Gaussian domain. The samples in the Gaussian domain can be discretely described by the inverse, Q^{-1} of the normal distribution

$$F(s_N) = Q^{-1} F(s) \tag{3}$$

or for individual samples

$$s_{\mathrm{NQT}}(t) = Q^{-1}(\Pr[s \le s(t)]) \tag{4}$$

where s_{NQT} is the Gaussian-transformed discharge or water level. Equation (4) describes the NQT of s. The same can be done for the error series, which results in a discrete population of errors in the Gaussian domain here denoted as e_{NQT}. To apply the inverse of the NQT, i.e. to convert any value in between the sampled points of s_{NQT} or e_{NQT} to associated values s and e, we use linear interpolation for points within the domain covered by the populations. If values are sought outside this domain linear extrapolation is applied on a number of points in the tails of the distribution. Equation (1) can now be applied in the normal domain,

$$s_{\mathrm{NQT}}(t) = \hat{s}_{\mathrm{NQT}}(t) + e_{\mathrm{NQT}}(t) \tag{5}$$

where the subscript "NQT" refers to variables that have been transformed into the Gaussian domain.

2.1.2 Quantile regression in the Gaussian domain

If one assumes that the residuals of a relation between errors and forecasted values (such as defined in Eq. (2).) are Gaussian, an estimate of the sample mean of the relation can be defined as the solution to the problem of minimizing a sum of squared residuals. This approach has been followed by Montanari and Brath (2004) who consequently estimate the variance of the relation to estimate effective uncertainty. If Gaussianity of the residuals of Eq. (2) cannot be assumed, one can turn to the estimation of the sample median and other quantiles instead. The latter approach has been chosen in this study by using Quantile Regression, which does not make any assumptions about the nature of the residuals of the forecast – error relationship. The Quantile Regression methodology is further explained below based on the Gaussian samples of e_{NQT} and s_{NQT}.

The sample median (as opposed to the sample mean) can be estimated by minimizing the sum of absolute residuals (as opposed to squared residuals). Minimizing a sum of asymmetrically weighted absolute residuals (by giving different weights to positive and negative residuals) can yield other (in fact, any) quantiles besides the median (for more details, see Koenker and Hallock, 2001). Applying this to the sampled values of s_{NQT} and e_{NQT} at a certain lead time of interest, this can be formulated as

$$\min \sum_{t=1}^{n} \rho_\tau \left[e_{\mathrm{NQT},\tau}(t) - \hat{e}_{\mathrm{NQT},\tau}\left(t, \hat{s}_{\mathrm{NQT}}(t)\right) \right] \tag{6}$$

with $\hat{e}_{\mathrm{NQT},\tau}$ the estimate of the Gaussian-transformed error at a certain quantile interval τ and ρ_τ the weighting function

that pushes $\hat{e}_{\mathrm{NQT},\tau}$ to its associated quantile location. In case $\tau = 0.5$ this would yield the unconditional median. Other values of τ can be used to determine other quantiles.

We assume that estimates of the conditional quantiles in the Gaussian domain can be described by the linear equation

$$\hat{e}_{\mathrm{NQT},\tau}(t) = a_\tau \hat{s}_{\mathrm{NQT}}(t) + b_\tau \quad (7)$$

where a_τ and b_τ are the regression constants, valid for the lead time of interest. After substitution of Eq. (7) into Eq. (6), a_τ and b_τ can be found efficiently by linear programming (Koenker, 2005; Koenker and Hallock, 2001).

2.1.3 Imposing the error models in operational forecasts

To describe the distribution of errors, conditional on the forecasted values, Quantile Regression programming functions within the R-package *quantreg* (Koenker, 2010) were used. To transform quantile estimates from the Gaussian domain to the untransformed (original) domain, linear interpolation has been used to connect the combinations of NQT transformed simulated values $\hat{s}_{\mathrm{NQT}}$ and estimated error quantiles $\hat{e}_{\mathrm{NQT},\tau}$ in the Gaussian domain, to their counterparts in the untransformed domain, using the following relationship:

$$\hat{s}_\tau(t) = \hat{s}(t) + \mathrm{NQT}^{-1}\left[a_\tau \hat{s}_{\mathrm{NQT}}(t) + b_\tau\right] \quad (8)$$

This yields calibrated discrete quantile relationships in the untransformed domain, which can be imposed on any forecasted value by means of linear interpolation or, if forecasted values are found outside the domain of the calibration dataset, with linear extrapolation.

The quantile error models can be derived at several lead times. The derived error models can consequently be applied in an operational context. An experimental module setup for imposing the error models in the standalone version of NFFS has been developed in the statistical computing language R (R Development Core Team, 2010). This R based module can be executed from within Delft-FEWS (Weerts et al., 2010a).

2.2 Case study descriptions

Figure 1 shows the locations of the case study areas relative to the coast line of England and Wales. The catchment in the North is the Upper Calder, the middle part is the Upper Severn with multiple interconnected catchments and the two catchments South are used in the Ravensbourne case study.

2.2.1 North East Region, Upper Calder, Todmorden (147 km^2)

The Upper Calder catchment drains an area of 147 km^2 to the river gauging station at Mytholmroyd, just upstream of Caldene Bridge. It is underlain by Carboniferous rocks of Millstone Grit and Coal Measures, with the former predominating in the high moorland areas. The river and its

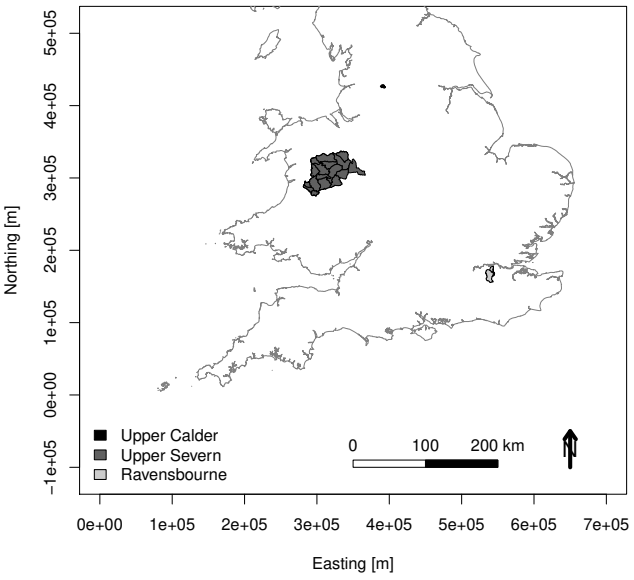

Fig. 1. Overview of the locations of the case study areas relative to the coastline of England and Wales. At the top the catchment of the Upper Calder, in the middle the Upper Severn with multiple interconnected catchments and at the bottom the two Ravensbourne catchments.

tributaries flow through steep and relatively narrow valleys. About 18% of the area (i.e. 26 km^2) drains to reservoirs. Typically these are reservoirs for direct water supply, releasing only compensation flows unless spilling at times of flood. The natural flow regime has also been modified by various channel improvements and flood defences, including schemes at Todmorden and Mytholmroyd. Upper Calder with forecast location Todmorden is part of the NFFS North-East Region (England, Environment Agency, NorthEast). The Upper Calder is a fast responding catchment and is modelled with a PDM rainfall runoff model (Moore, 2007). The PDM model is forced by the input created by a snow model.

2.2.2 Midlands Region, Upper Severn, various locations and basin sizes (150–1000 km^2)

The River Severn rises in the Cambrian mountains at Plynlimon at a height of 741 m AOD and flows in northeasterly direction through Llanidloes, Newtown and Welshpool before meeting the Vyrnwy tributary upstream of Shrewsbury. The valley is wide and flat in this confluence area, with a considerable extent of floodplain.

The river then flows through Montford to Shrewsbury, and is joined at Montford Bridge by the River Perry which flows from the Oswestry area to the North. The lowest point in the Upper Severn catchment is defined by Midlands Region as the gauge at Welshbridge in Shrewsbury. There are also significant areas of floodplain in the reach from Shrawardine, upstream from Montford, to Welshbridge. Average annual

rainfall can exceed 2500 mm in the Cambrian mountains in the upper reaches of the Severn, and in Snowdonia National Park, in the upper reaches of the Vyrnwy catchment. The catchment area to Welshbridge is approximately 2284 km^2 including 778 km^2 for the Vyrnwy catchment to the flow gauge at Llanymynech. The main reservoirs in the Upper Severn catchment are Lake Vyrnwy in the upper reaches of the Vyrnwy, and the Clywedog Reservoir in the upper reaches of the Severn. There are no significant flow control structures downstream of Welshbridge in the remainder of the Upper Severn catchment. The Upper Severn catchment is represented in NFFS by a combination of MCRM rainfall-runoff models (Bailey and Dobson, 1981; Wallingford Water, 1994) and DODO routing models (Wallingford Water, 1994).

2.2.3 Thames Region, Ravensbourne, two locations (32 and 68 km^2)

The River Ravensbourne and tributaries drain highly urbanised areas from the south of London towards Lewisham before joining the Thames at Deptford Creek. The total Ravensbourne catchment area equals approximately 180 km^2. The more slowly responding rural part of the Ravensbourne South Branch catchment makes up around 55 km^2, which does not contribute significantly to flood events and is therefore generally discarded from analyses of flood hydrology (as verified by historic calibration data over many events).

The remaining 125 km^2 is highly urbanised and has a very rapid response to rainfall and, due to the large proportion of paved surfaces, there is very little hydrological memory (antecedent storage). As a result, hydrographs throughout the catchment often rise steeply from baseflow to threshold levels in around 30 min (sometimes even less), and fall again almost as rapidly at the upstream locations, whilst locations lower in the catchment take slightly longer to recede (due to later arrival of upstream contributions). The NFFS integrated catchment model for Ravensbourne (Thames Region) comprises 16 TCM models (Greenfield, 1984; Wilby et al., 1994) providing inputs to an ISIS hydrodynamic model.

2.3 Data used in the case studies

For both the calibration and validation only operationally available data are used. The data available from the archive consist of RTS data (observed level data, rain gauge data, air temperature and catchment average rainfall data), Radar Actuals, Radar Forecast, and Numerical Weather Prediction data (all three from UK Met Office) and is available from 2006 onwards. The operational forecasts of river discharges and water levels benefit from the availability of (near) real-time observations. These are propagated into the model by means of data assimilation, prior to forecasting.

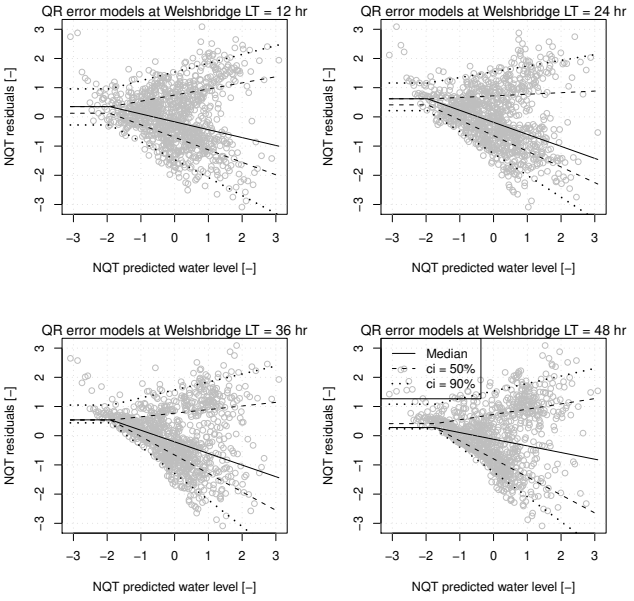

Fig. 2. Transformed forecast errors versus transformed forecasted water levels (both in Gaussian domain) together with the derived Quantile Regressions for the different confidence levels.

The data was split in a calibration data set (2006 and 2007) and a validation data set (2008 and 2009). Both the calibration and validation datasets contain several major flood events. Unfortunately, this is not the case for the validation data set for the Thames region as, here, only few events were observed in 2008 and 2009.

3 Results and discussion

3.1 Derivation of forecast error models

The Quantile Regression methodology to describe the predictive uncertainty as described in Sect. 2 has been applied to water levels at several forecast locations in the National Flood Forecasting System (NFFS). In the calibration phase, the forecast error models are derived. Figure 2 shows an example of the calibration for the Upper Severn forecast location Welshbridge (2077) in the Gaussian domain. Figure 3 shows the example of the derived forecast error model in the untransformed domain. The effect of the NQT^{-1} on the derived 50% and 90% quantiles is evident: the fitted Quantile Regression relationships change from linear to nonlinear under influence of the back-transformation, while describing the heteroscedastic behaviour of the forecast errors in water levels as function of predicted water level. When fitting the different quantiles during calibration, sometimes the quantiles were found to cross at low water levels. This is an artefact of the fitting procedure and this problem has been overcome by using a fixed error model below certain water levels in the Gaussian domain as is visible in Fig. 2. This fixed

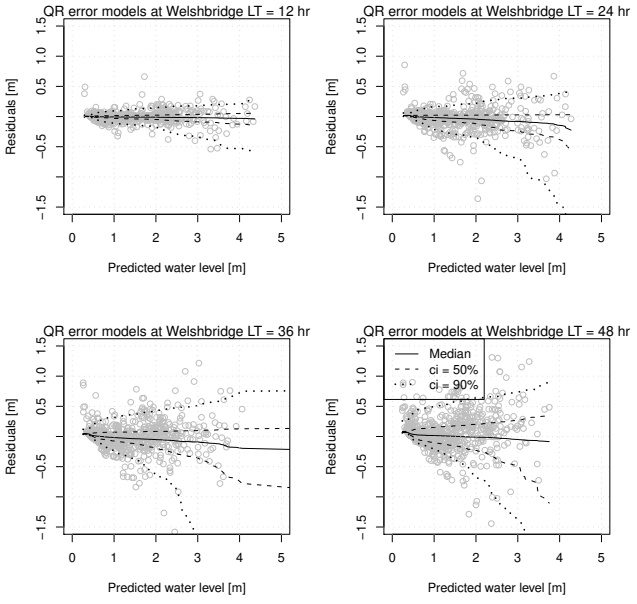

Fig. 3. Forecast errors versus forecasted water levels together with the derived Quantile Regressions for the different confidence levels.

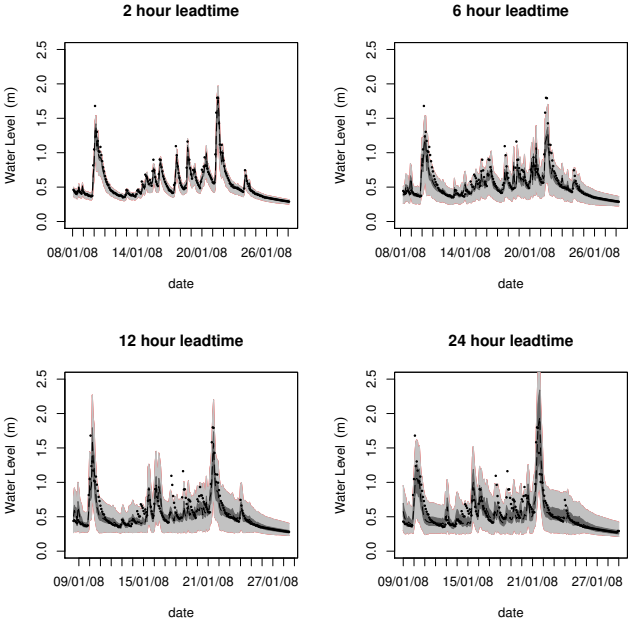

Fig. 4. Validation of Quantile Regression method for water level forecasts at Todmorden for the January 2008 events for 2, 6, 12, and 24 h leadtimes. The dark grey area represents the 50% confidence interval and the light grey area represents the 90% confidence interval, the black dashed line the 50% estimate, and the black dots the observations.

error model is hardly visible after the NQT^{-1} (see Fig. 3), because it describes only a small portion of the full flow regime domain.

The effect of fitting the quantiles on forecast values – instead of on forecast errors – was also investigated. We found (not shown) that the regressions based on forecast errors produced more reliable results, while the regression based on forecast values resulted in more unreliable results (especially in the high water level domain).

The advantage of doing the Quantile Regression in the Gaussian domain is that no subjective assumptions on the regression have to be made, which limits the number of parameters to be fitted – 2 per quantile, see Eq. (7). Other approaches may be even more parsimonious than the approach taken here, such as for instance the Model Conditional Processor (MCP, Todini, 2008), which requires the estimation of 1 parameter in order to determine the full density of the predictive uncertainty. This method, however, assumes that the residuals of the forecast – error relationship are normally distributed. Making this assumption is not necessary using the approach described here.

After deriving these distributions for all forecast locations and at all lead times used in the case studies, they can subsequently be used in NFFS in real-time or hindcast mode. For all case studies, the 5%, 25%, 50%, 75%, 95% quantiles were derived and used for validation.

3.2 Validation of forecast error models

Below, the application of Quantile Regression to provide a probabilistic forecast is presented for the three case studies, representing a wide variety of hydrological conditions and catchment sizes.

3.2.1 North East Region, Upper Calder

Figure 4 shows the validation of the Quantile Regression method for the January 2008 events for 4 leadtimes (2, 6, 12 and 24 h). Most of the time, the observations fall within the 90% predictive confidence interval. It is also clear that the uncertainty increases with leadtimes as was to be expected. One can also see that the uncertainty increases and decreases depending on the forecasted water level and lead time. In Fig. 4, it is also visible that some observed values lie outside the confidence interval. For a 90% confidence interval, and for each forecasted water level, this should happen in only 10% of all cases. Because the sample size at high forecasted and observed water level in both the calibration and validation phase is small (only few major and minor events, see for example Fig. 2) it is very difficult to make strong statements about the performance of the Quantile Regression method at higher water levels. Unfortunately, this is the area of main interest.

Figure 5 shows the validation results for the January 2009 events. The forecasted water levels for January 2009 for longer leadtimes are not as good as in January 2008 (as can be seen by the 50% estimate). There are several events forecasted that did not occur especially at longer leadtimes.

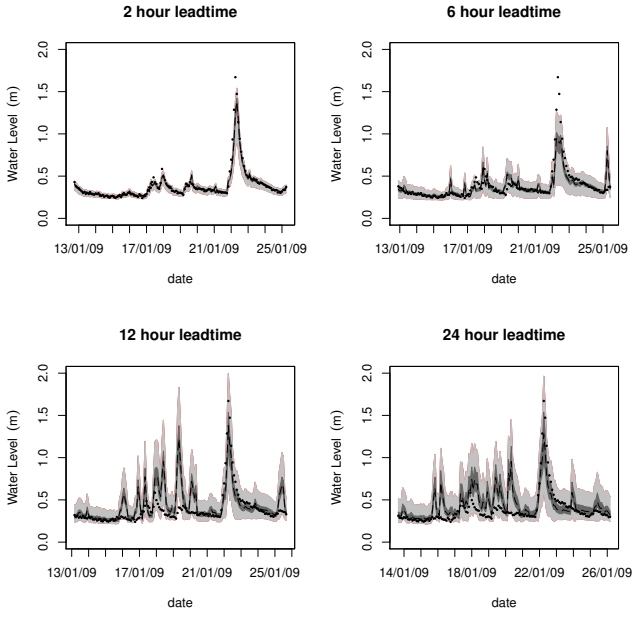

Fig. 5. Validation of Quantile Regression method for water level forecasts at Todmorden for the January 2009 events for 2, 6, 12 and 24 h leadtimes. The dark grey area represents the 50% confidence interval and the light grey area represents the 90% confidence interval, the black dashed line the 50% estimate, and the black dots the observations.

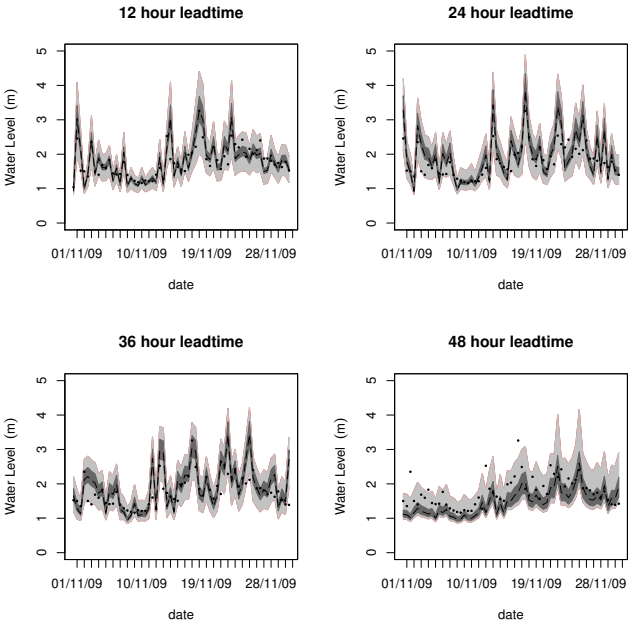

Fig. 6. Validation of Quantile Regression method for water level forecasts at Welshbridge (2077) for the November 2009 events for 12, 24, 36 and 48 h leadtimes. The dark grey area represents the 50% confidence interval and the light grey area represents the 90% confidence interval, the black dashed line the 50% estimate, and the black dots the observations.

Table 1. Percentage of observations within respective predictive confidence intervals for period January 2008–July 2009 for forecast location Todmorden (North East Region).

Location id & name	predictive confidence interval	2 h	4 h	6 h	12 h	18 h	24 h
TODMDN1	25%–75%	61.8	65.5	66.4	48.0	41.5	39.9
Todmorden	5%–95%	89.2	90.2	90.6	89.7	90.1	90.4

The reason for this is not yet clear. The wider confidence bounds at the peak values also indicate that this behaviour, to some degree, was present in the calibration data set. However, it is also possible that it is related to changes over time (i.e. between calibration and validation periods) in numerical weather prediction products used for the forecast.

Both Figs. 4 and 5 show that the Quantile Regression method tested here can not correct "bad" forecasts. Generally, the confidence intervals are accurately estimated even though they were only estimated from a hindcast covering a short period (2006–2007). The Quantile Regression method closely resembles the confidence interval wherein one expects the observation. The performance is further illustrated by Table 1 which shows the percentage of the observations within the confidence intervals at various lead times.

3.2.2 Midlands Region, Upper Severn

The case study on the Upper Severn using Quantile Regression is focused on several locations (see Figs. 1 and 7). These locations consist of both upstream (modelled with MCRM) and downstream (modelled using DODO) forecast locations and were chosen to show how Quantile Regression can be used to derive predictive confidence intervals in an end-to-end forecasting system. Figure 6 shows the validation for the flood events of November 2009 at Welshbridge (2077) at 12, 24, 36 and 48 h lead time. This figure shows how Quantile Regression can give estimates of the confidence intervals during flood events. Validation results for the other locations are given in Fig. 7 as Quantile-Quantile (Q-Q) plots. Most Q-Q plots follow the 1:1 line closely. Figure 7 also shows poor performance at Caersws and to some degree at Welshpool. The reason for the poor performance at Caerws is the combination of the relatively large bias in the simulation of the DODO model for Caerws and the internal error correction procedure that is applied for this location. This internal error correction procedure causes the forecast at Caerws to jump, often one or two hours into the forecast, from the error corrected foreast to the biased forecast. This happens more often in the validation than in the calibration period, therefore when switching the validation with the calibration period this results in exactly the opposite figure (all lines above 1:1 line)

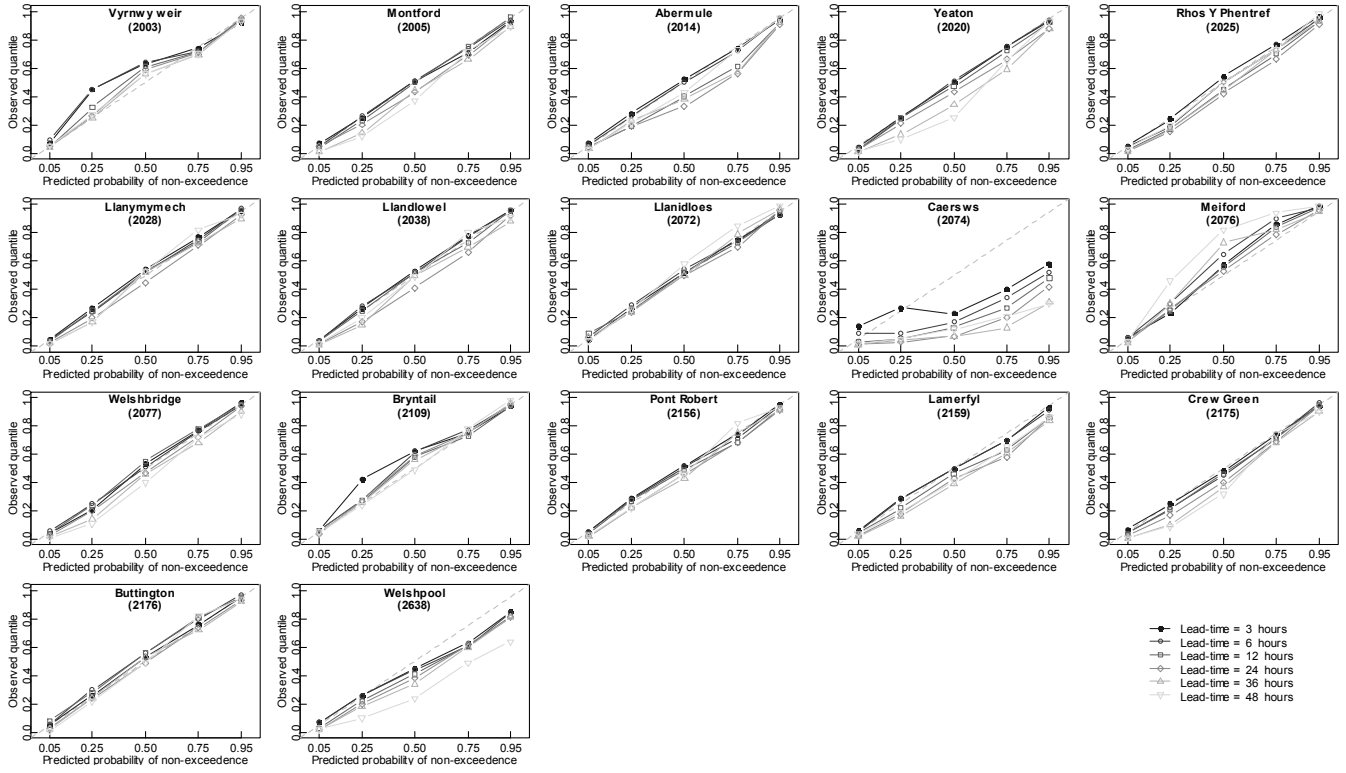

Fig. 7. Reliability Q-Q plots (predicted non-exceedence probabilities versus the fraction of observations that are lower than the corresponding estimated value indicated by "observed quantiles") for all Upper Severn locations for all lead times considered for the validation period (2008 and 2009). Bracketed numbers are unique station identifiers.

Table 2. Percentage of observations within respective predictive confidence intervals for period January 2008–July 2009 for forecast locations Ravensbourne (Thames Region).

Location id & name	predictive confidence interval	3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h
3470TH	25%–75%	56.4	51.1	44.4	41.2	42.5	42.5	42.8	56.4
Catford Hill	5%–95%	96.7	96.2	95.5	95.4	94.3	93.5	92.3	96.7
3489TH	25%–75%	65.1	58.7	56.2	55.4	55.7	55.1	55.0	65.1
Manor House Gardens	5%–95%	96.0	94.8	94.5	94.4	94	93.8	94.2	96

for Caerws (not shown). To improve the performance at Caerws the DODO model and the error correction procedure should be improved.

3.2.3 Thames Region, Ravensbourne

Figure 8 shows the validation results of the method for the February 2009 event for Manor House Gardens (3489TH). Most of the time the observations fall within the 90% confidence interval. It is also clear that the uncertainty increases with lead times as expected. One can also see that the uncertainty increases and decreases depending on the forecasted level. The performance is further illustrated by Table 2 which shows the percentage of the observations within the confidence intervals at various lead times. Table 2 shows that for these two locations the derived forecast error models are somewhat too wide (underconfident) as too many observations lie within the 50% and 90% confidence limits. However, this could be resulting from the fact that the period 2008–2009 was dissimilar to the calibration period 2006–2007. In 2008 and 2009, there were relatively few small, medium and major flood events for the whole of the Thames region. This should be further tested with a longer record that corresponds more closely with the calibration period.

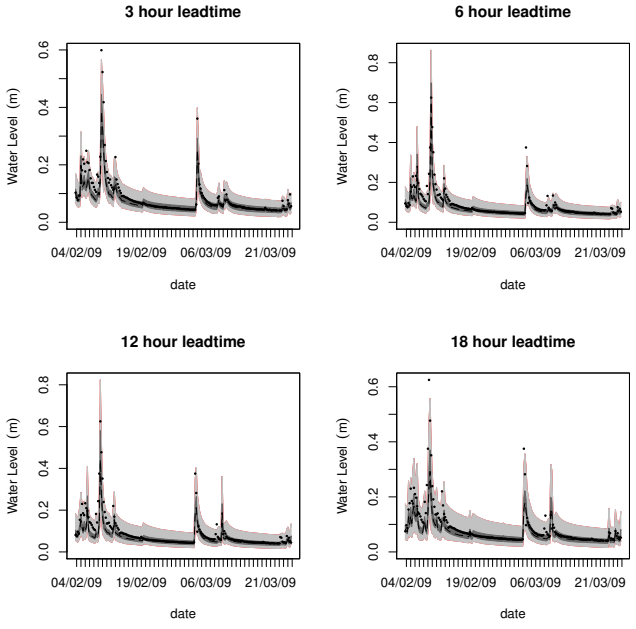

Fig. 8. Validation of Quantile Regression method for water level forecasts at Manor House Gardens for the February 2009 event for 3, 6, 12 and 18 h lead times. The dark grey area represents the 50% confidence interval and the light grey area represents the 90% confidence interval, the black dashed line the 50% estimate, and the black dots the observations.

4 Conclusions

A method to provide predictive uncertainty estimates of water level or flow forecasts is presented. The method aims to characterise the distribution of the water level or flow forecast error conditioned on the value of the predicted water level or flow by means of Quantile Regression relationships at quantiles of interest. The method does not consider the independent sources of uncertainty but instead considers the effective uncertainty of the forecast process only, which can be a result of input or output uncertainty, model structural uncertainty or parameter uncertainty. The quantile error relationships are estimated in the Gaussian domain. To this end, both the available forecast population and error population are made Gaussian by means of a Normal Quantile Transform. Several sets of Quantile Regression relationships may be derived at specific lead times that are of interest to the user. The Quantile Regression relationships can straightforwardly be implemented as a post-processor in a real-time, operational forecasting system.

The method was tested by deriving Quantile Regression relations for several lead times using a calibration hindcast set and consequently predicting forecast errors of water levels using an independent validation set in three case studies in a stand alone version of the National Flood Forecasting System of England and Wales. The three case studies across England and Wales contain a variety of catchments (size and behaviour) and hydrological models. From the results, we can conclude that the Quantile Regression method to provide predictive uncertainty estimates of water level forecasts is promising. We showed that the derived Quantile Regression relationships predict the error quantiles satisfactorily at all lead times. Moreover, the developed method is simple to apply, requires very few assumptions and is easy to understand by both scientists and forecasters.

However, there are also some limitations. The method, like all post-processing methods, requires a long (homogeneous) calibration and validation set. In the case studies, only short calibration and validation sets were available, containing only few extreme events. This compromised error descriptions at higher predicted water levels, which could only be made through extrapolation of the derived quantile regressed relations. The case study containing a validation period without extreme events (e.g. Ravensbourne case study) indicated that currently available data records may cause problems with deriving and testing the method.

Furthermore, homogeneity of data sets can be an issue (observed in particular in the Upper Calder case study). Such inhomogeneities may be caused by a number of factors, such as (a) changes in the hydraulics/hydrology of the river/catchment changes considerably (in fact, in this case the hydrological/hydraulic model itself needs re-calibration) and (b) changes in the forecasting system (e.g. by adaptation of state updating procedures or models used) or (c) changes in external data sources used to drive the forecasting system (e.g. meteorological models or external observed boundary conditions). Like any other statistical post-processing method, the developed method requires recalibration if such inhomogeneities occur.

If these limitations are considered by the user, the Quantile Regression method can be straightforwardly employed in operational forecasting because (a) the required data is always available in the operational context and (b) the methodology requires little computation time.

Acknowledgements. The authors would like to acknowledge Andrew Wood (NOAA-NWS) for bringing Quantile Regression to our attention during the June 2008 HEPEX workshop in Delft on hydrological ensemble post-processing methods. The authors would also like to thank R. Koenker, E. Todini and an anonymous reviewer who helped to improve our manuscript considerably. This work described in this paper was carried out under R&D project SC080030 "risk based probabilistic flood forecasting for integrated catchment models" under the joint Environment Agency/Defra Flood and Coastal Risk Management R&D programme.

Edited by: F. Pappenberger

References

Bailey, R. and Dobson, C.: Forecasting for floods in the Severn catchment, J. Inst. Water Engrs. Sci., 35(2), 168–178, 1981.

Baur, D., Saisana, M., and Schulze, N.: Modelling the effects of meteorological variables on ozone concentration – a quantile regression approach, Atmos. Environ., 38(28), 4689–4699, doi:10.1016/j.atmosenv.2004.05.028, 2004.

Beven, K. J.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, 2006.

Bremnes, J.: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output, Mon. Weather Rev., 132(1), 338–347, 2004.

Bremnes, J.: A comparison of a few statistical models for making quantile wind power forecasts, Wind Energy, 9(1–2), 3–11, doi:10.1002/we.182, 2006.

Broersen, P. M. T. and Weerts, A. H.: Automatic error correction of rainfall-runoff models in flood forecasting systems, in: Instrumentation and Measurement Technology Conference (IMTC), 2005, Proceedings of the IEEE, 19 May 2005, vol. 2, 963–968, 2005.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model, Adv. Water Resour., 31(10), 1309–1324, doi:10.1016/j.advwatres.2008.06.005, 2008.

Greenfield, B.: The Thames Water Catchment Model, Internal report, Technology and Development Divisionbn, Thames Water, UK, 1984.

Juban, J., Siebert, N., and Kariniotakis, G.: Probabilistic short-term wind power forecasting for the optimal management of wind generation, 2007 IEEE Lausanne Powertech, Vols. 1–5, 683–688, 2007.

Kelly, K. S. and Krzysztofowicz, R.: A bivariate meta-Gaussian density for use in hydrology, Stoch. Hydrol. Hydraul., 11(1), 17–31, 1997.

Koenker, R.: Quantile Regression, Cambridge University Press, 2005.

Koenker, R.: Quantile regression in R: A vignette, online available from: http://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf, 2010.

Koenker, R. and Basset, G.: Regression Quantiles, Econometrica, 46(1), 33–50, 1978.

Koenker, R. and Hallock, K. F.: Quantile Regression, The Journal of Economic Perspectives, 15(4), 143–156, 2001.

Krzysztofowicz, R. and Maranzano, C. J.: Hydrologic uncertainty processor for probabilistic stage transition forecasting, J. Hydrol., 293(1–4), 57–73, doi:10.1016/j.jhydrol.2004.01.003, 2004.

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.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483–499, doi:10.5194/hess-11-483-2007, 2007.

Nielsen, H., Nielsen, T., Madsen, H., Giebel, G., Badger, J., Landberg, L., Sattler, K., Voulund, L., and Tofting, J.: From wind ensembles to probabilistic information about future wind power production – results from an actual application, 2006 International conference on probabilistic methods applied to power, 639–646, 2006.

Pitt, M.: The Pitt Review – Learning Lessons from the 2007 floods, online available from: http://archive.cabinetoffice.gov.uk/pittreview/thepittreview.html, last access: 9 June 2010, 2008.

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, online available from: http://www.R-project.org, 2010.

Raferty, A., Balabdaoui, F., Gneiting, T., and Polakowski, M.: Using bayesian model averaging to calibrate forecast ensembles, University of Washington, Department of Statistics, Seattle, Washington, 2003.

Raiffa, H. and Schlaifer, R.: Applied Statistical Decision Theory, 1e ed., Wiley-Interscience., 2000.

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(1–2), 56–69, doi:10.1016/j.jhydrol.2008.03.027, 2008.

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, Wat. Resour. Res., 45, doi:10.1029/2007WR006758, 2009.

Seo, D.-J., Herr, H. D., and Schaake, J. C.: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, Hydrol. Earth Syst. Sci. Discuss., 3, 1987–2035, doi:10.5194/hessd-3-1987-2006, 2006.

Seo, D. J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, J. Hydrol., 367(3–4), 255–275, doi:10.1016/j.jhydrol.2009.01.019, 2009.

Serban, P. and Askew, A. J.: Hydrological forecasting and updating procedures, IAHS pub. no. 201, 357–369, 1991.

Shamseldin, A. and O'Connor, K.: A non-linear neural network technique for updating of river flow forecasts, Hydrology and Earth System Sciences, 5(4), 577–597, 2001.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C.: Probabilistic quantitative precipitation forecasting using Bayesian model averaging, Monthly Weather Review, 135(9), 3209–3220, doi:10.1175/MWR3441.1, 2007.

Sousa, S. I. V., Pires, J. C. M., Martins, F. G., Pereira, M. C., and Alvim-Ferraz, M. C. M.: Potentialities of quantile regression to predict ozone concentrations, Environmetrics, 20(2), 147–158, doi:10.1002/env.916, 2009.

Todini, E.: Hydrological catchment modelling: past, present and future, Hydrol. Earth Syst. Sci., 11, 468–482, doi:10.5194/hess-11-468-2007, 2007.

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Intl. J. River Basin Management, 6(2), 123–137, 2008.

van der Waerden, B.: Order tests for two-sample problem and their power I, Indagat. Math., 14, 453–458, 1952.

van der Waerden, B.: Order tests for two-sample problem and their power II, Indagat. Math., 15, 303–310, 1953a.

van der Waerden, B.: Order tests for two-sample problem and their power III, Indagat. Math., 15, 311–316, 1953b.

Wallingford Water: A flood forecasting and warning system for the river Soar, Wallingford Water, Wallingford, 1994.

Weerts, A. H., Schellekens, J., and Sperna Weiland, F.: Real-

Time Geospatial Data Handling and Forecasting: Examples From Delft-FEWS Forecasting Platform/System, IEEE JSTARS, doi:10.1109/JSTARS.2010.2046882, 2010a.

Weerts, A. H., El Serafy, G. Y., Hummel, S., Dhondia, J., and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Comput. Geosci., 36(4), 453–463, doi:10.1016/j.cageo.2009.07.009, 2010b.

Werner, M., van Dijk, M., and Schellekens, J.: DELFT-FEWS: an open shell flood forecasting system, in Proceedings of the 6th International Conference on Hydroinformatics, 1205–1212, World Scientific Publishing Company, 2004.

Werner, M., Cranston, M., Harrison, T., Whitfield, D., and Schellekens, J.: Recent developments in operational flood forecasting in England, Wales and Scotland, Meteorological Applications, 16(1), 13–22, doi:10.1002/met.124, 2009.

Wilby, R., Greenfield, B., and Glenny, C.: A coupled synoptic-hydrological model for climate change impact assessment, Journal of Hydrology, 153(1–4), 265–290, doi:10.1016/0022-1694(94)90195-3, 1994.

Wood, A. and Schaake, J.: Correcting errors in streamflow forecast ensemble mean and spread, J. Hydrometeorol., 9(1), 132–148, doi:10.1175/2007JHM862.1, 2008.