Hydrology and Earth System Sciences Discussions

Technical Note: The Normal Quantile Transformation and its application in a flood forecasting system

K. Bogner[1], F. Pappenberger[2], and H. L. Cloke[3]

[1] Institute for Environment and Sustainability, European Commission Joint Research Centre, Via E. Fermi, 2749, 21027 Ispra (VA), Italy
[2] European Centre for Medium-Range Weather Forecasts Shinfield Park Reading, RG2 9AX, UK
[3] Department of Geography, King's College London, London, UK

Correspondence to: K. Bogner (konrad.bogner@jrc.ec.europa.eu)

	Discussion Paper
	Discussion Paper
	Discussion Paper
	Discussion Paper

HESSD

8, 9275–9297, 2011

Application of the NQT

K. Bogner et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄◄	►►
◄	►
Back	Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

The Normal Quantile Transform (NQT) has been used in many hydrological and meteorological applications in order to make the Cumulated Density Function (CDF) of the observed, simulated and forecast river discharge, water level or precipitation data
5 Gaussian. It is also the heart of the meta-Gaussian model for assessing the total predictive uncertainty of the Hydrological Uncertainty Processor (HUP) developed by Krzysztofowicz. In the field of geo-statistics this transformation is better known as Normal-Score Transform. In this paper some possible problems caused by small sample sizes for the applicability in flood forecasting systems will be discussed and illus-
10 trated by examples. For the practical implementation commands and examples from the freely available and widely used statistical computing language R (R Development Core Team, 2011) will be given (represented in `Courier` font) and possible solutions are suggested by combining extreme value analysis and non-parametric regression methods.

1 Introduction

The Normal Score Transform or NQT has been applied in various fields of geoscience in order to make the mostly asymmetrical distributed real world observed variables more treatable and to fulfil the basic underlying assumption of normality, which is intrinsic to most statistical models (e.g. Moran, 1970; Goovaerts, 1997; Deutsch and
20 Journel, 1998). The meta-Gaussian family of distribution functions (Kelly and Krzysztofowicz, 1997) allows the marginal distribution functions of the variates to take any form and the dependence structure between any two variates to be monotone nonlinear and heteroscedastic. This most convenient property has been incorporated into the HUP (Krzysztofowicz and Kelly, 2000; Krzysztofowicz and Herr, 2001; Krzysztofowicz
25 and Maranzano, 2004), which is now part of several operational forecasting systems (e.g. Todini, 2008; Reggiani et al., 2009; Bogner and Pappenberger, 2011) in order to estimate the predictive uncertainty of the hydrological forecasts.

In Krzysztofowicz (1997) the theory behind the NQT is outlined and the practical application is demonstrated e.g. in Montanari (2005); Seo et al. (2006). The main objective of this study is to show the difficulties occurring in the inversion of the empirical NQT, if the normal random deviates lie outside the range of the historically observed range.

Following the work of Krzysztofowicz (1997) the empirical NQT involves the following steps (in brackets the corresponding R commands):

1. Sorting the sample X from the smallest to the largest observation, $x_{(1)}, \ldots, x_{(n)}$ (`x=sort(x)`)

2. Estimating the cumulative probabilities $p_{(i)}, \ldots, p_{(n)}$ using a plotting position like $i/(n+1)$ such that $p_{(i)} = P(X \leq x_{(i)})$ (`p=ppoints(x,a)`, with $a = 0$ for the Weibull, resp. $a = 1/2$ for the Hazen plotting position)

3. Transforming each observation $x_{(i)}$ of X into observation $y_{(i)} = Q^{-1}(p_{(i)})$ (`y=qnorm(p)`) of the standard normal variate Y, with Q denoting the standard normal distribution and Q^{-1} its inverse, applying discrete mapping (`f=approx(x,y)`)

The first two steps (`qnorm(ppoints(x))`) are integrated in the command `qqnorm(x)`.

The problems of applying the NQT arise for the reverse process, when the sampled data points in the normal space fall outside the range of historical samples (i.e. probability quantiles greater than $n/(n+1)$). In order to be able to extrapolate to extreme values, which are rarely observed in the historical samples due to the limited amount of available data, different parametric and non-parametric approaches have been tested in this paper. The problem of a sufficient amount of data is naturally very common for example in flood frequency estimation (Laio et al., 2009), downscaling of climate change scenarios for hydrological applications (Bo et al., 2007; Vrac and Naveau, 2007, 2008), and hydrology in general (Zhu, 1987; Engeland et al., 2004).

This paper will specifically concentrate on the impact of small sample sizes in real-time flood forecasting using ensemble driven systems in combination with the HUP.

In the next section a forecast example of the European Flood Alert System (EFAS) is shown in order to demonstrate the problem of the back-transformation and its impact on the predictive uncertainty. Then several different solutions are given and some advantages and disadvantages outlined. Finally some concluding remarks and practical advice is given.

2 Example forecast

The EFAS (Thielen et al., 2009; Bartholmes et al., 2009) produces daily stream-flow forecasts and includes post-processing through data assimilation and error correction at selected runoff gauging stations.

Since 2010 ensembles of stream-flow forecasts issued daily are corrected and predictive uncertainties are estimated at some stations, for which historical time-series of simulations and observations are available as well as observations in real-time (see for details Bogner and Pappenberger, 2011). In Table 1 characteristic values from two selected stations are given, Bohumin (Odra river, CZ) and Hofkirchen (Danube, DE), which will be analysed in more detail. For the calibration of the post-processor, time series of daily observed and simulated stream-flow data are necessary, which can be quite different in length and occurrences of floods at the various stations investigated. For example the data at station Bohumin comprises six years only with a maximum observed discharge of less than $600\,\mathrm{m^3\,s^{-1}}$, which corresponds roughly to a flood event with a return period between two and five years, whereas at Hofkirchen eight years of data are available including some severe flood events with twenty to fifty years return periods (see Fig. 1). In 2010 during the testing period for the EFAS postprocessor, the forecasted discharge at Bohumin far exceeded the maximum of the historical data sample, and this led to the initiation of this study. Consequently we present the case of Bohumin here. At station Hofkirchen the effect of sample size on the extrapolation

methods will be analysed in more detail by comparing the results of the total available and a split sample (divided into two halves).

The post-processor runs operationally twice a day and includes the minimization of the error between the most recent past observed and simulated discharge values and the correction of the deterministic ten days ahead forecasts and the corresponding forecasts derived from two different Ensemble Prediction Systems (EPS). The NQT is applied prior to the post-processing to all available stream-flow data (measured, simulated and predicted) and the results of the error corrected forecasts are transformed back from normal space into the real world and the predictive uncertainty is estimated. In Fig. 2 the final output of the post-processor is shown with forecast initiation at time-step zero (dashed vertical line). The past eight days $(-8, \cdots, -1)$ are included to demonstrate the performance of the error correction showing the corrected one step ahead predictions, the observations and the prediction uncertainties estimated by the HUP (Krzysztofowicz and Kelly, 2000). From lead-time one onwards $(1, \cdots, 10)$ the error corrected forecasts are shown including two stream-flow forecasts based on deterministic weather forecast systems (DWD, ECMWF-det.) and two ensemble prediction systems (51 members EPS from ECMWF and 16 members COSMO-LEPS). The resulting total predictive uncertainty integrating the model input uncertainty (i.e the weather forecast uncertainty) and the hydrological uncertainty is calculated for the different lead-times and is shown as shaded areas. Additionally two thresholds are indicated, the MQ value (lower horizontal line) representing the mean daily average discharge and the MHQ (upper horizontal line) representing the daily mean annual maximum discharge.

In Fig. 2a,d the problem of back-transforming data values exceeding the maximum of the historical sample used for calibration is demonstrated. For the observations corresponding to the one step ahead predictions, the NQT has been applied and the back-transformed measurements can show discrepancies to the real observations, because of the upper limit (observed maximum) in the historical data set. That is the reason why the observed values in Fig. 2a at time-step zero (forecast initiation) and leadtime one do not match the real observed data, but correspond to the historical observed

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

maximum. This limitation clearly indicates the necessity for including methods for extrapolation. It should be noted that this problem of extrapolation will only occur in the case of applying empirical NQT's, that means transforming empirical CDF's, and could be circumvented by fitting a theoretical CDF to the historical data sets. However the fitting of theoretical distribution functions is quite difficult under non stationary conditions and for data sets showing long-range dependencies, what is typically the case for hydrological time-series. Additionally for different regions probably different distribution functions will be optimal, which would make a supervised fitting and detailed verification at each single station (i.e. stream-flow pixel in spatially distributed models) nearly impossible, especially for continental scale forecast systems such as EFAS which runs on a 5 km grid over the whole of Europe.

3   Extrapolation methods

In this paper we concentrate on two approaches which represent a large class of possibilities and allow us to evaluate them in a flood forecasting specific setting:

(1) The first method is based on extreme value theory and tries to estimate future possible extreme stream-flow values by fitting a theoretical distribution to the upper (and lower) tail of the sample. The resulting extreme values are combined with the historical sample in order to find an optimal transform function. There are multiple approaches which could be used for extrapolation based on extreme value theory, for example: normal and log-normal distributions and 3-parameter log-normal; Log-Pearson Type III; Extreme Value type I, II, or III; Generalized Extreme Value (GEV); Logistic and General logistic; Goodrich/Weibull distribution; Exponential distribution; or Generalized Pareto Distribution (GPD) – to name a few. More mathematical and statistical details concerning extreme value theory can be found for example in Coles (2001); Finkenstädt and Rootzén (2004).

(2) In contrast to this approach the second is a non-parametric regression method called Generalized Additive Model (GAM, Hastie and Tibshirani, 1986), where the

regression (i.e. transformation) function is estimated directly without specifying its parametric form explicitly. GAMs are Generalized Linear Models (GLMs, McCullagh and Nelder, 1989) in which the linear predictor is specified partly in terms of a sum of smooth functions of covariates and have been introduced for modelling non-linear relationships (Hastie and Tibshirani, 1990; Wood, 2000, 2006). Generalized additive mixed models have been proposed for over-dispersed and correlated data, which arise frequently in hydrology (Lin and Zhang, 1999) and that gave the reason also for choosing the GAM in this study. Many more different non-linear (regression) models exist, such as neural networks, non-linear prediction methods (e.g. Laio et al., 2003) and kernel based support vector machines (e.g. Yu et al., 2006), but their application is beyond the scope of this technical note. In Fig. 3 some non-linear relationships between the standardized normal observations and the observations are shown demonstrating the appropriateness of GAMs in fitting this non-linear transformation function.

3.1 Extreme values

The application of extreme value theory in hydrology has a long tradition and an associated large literature. Fisher and Tippett (1928) started to work on the asymptotic theory of extremes, whereas in Gnedenko and Kolmogorov (1949/68) the theory for independent identically distributed random variables was completed. Fitting methods of extreme value type distributions to reliability data are outlined thoroughly in the famous work of Gumbel (1958). The Gumbel distribution is frequently applied, which is a double exponential distribution representing the limiting distribution for Gaussian data. Recently the GEV for annual maxima series (e.g. Ailliot et al., 2011) and the GPD for maxima exceeding thresholds have found the most attraction in environmental extreme value analysis (e.g. MacKay et al., 2011; Moloney and Davidsen, 2011; Mazas and Hamm, 2011).

For the practical implementation in R several packages are available, like ismev, evir and POT (Ribatet, 2006), which has been used in this study. In the Peaks Over Threshold (POT) model the limiting distribution of normalised excesses over

a threshold converges to the GPD, as the threshold approaches the endpoint (Pickands, 1975; Davison and Smith, 1990).

In the `POT` package different fitting methods are implemented ranging from methods of moments, unbiased probability weighted moments to maximum likelihood (`mle`, Ashkar and Tatsambon, 2007) and maximum goodness-of-fit (`mgf`, Luceño, 2006) estimators. Although Hosking and Wallis (1987) recommended the method of probability weighted moments for small sample sizes, the following two-step approach has been applied here. At first the GPD parameter were estimated by minimizing the Kolmogorov–Smirnov (`KS`) goodness-of-fit statistics, which were taken in step two as initial values for optimizing the maximum likelihood function by the use of the `Nelder--Mead` method (Nelder and Mead, 1965) resulting in stable parameter estimates and good agreements between the fitted and the empirical maxima (Fig. 4). The disadvantage of the POT model is the somewhat subjective choice of a threshold u and the necessity to de-cluster the possible serial correlated time-series by defining some criterion for making the observed events independent (i.e. defining the minimum timespan between two consecutive events not exceeding the threshold, see also Bogner et al., 2011, for more details). In the following example of R commands the observed data series x is de-clustered first by defining a timespan of eight days to ensure independence of events (see also Fig. 1).

```
require(POT)
events ← clust(x,u=u,tim.cond=8/365)
mgf.fit ← fitgpd(events,thresh=u,est="mgf",stat="KS")
mle.fit ← fitgpd(events,thresh=u,est="mle",method="Nelder-Mead",
control=list(start=list(scale=mgf.fit$para[1],shape=mgf.fit$para[2])))
```

3.2 Nonparametric regression

One remarkable property of the GAM is it's flexibility permitting the data to influence the shape of the smooth functions used for exploring the relationship between the transformed variable (response) and the observation (explanatory variable). In allowing user

defined penalization in the fitting by the use of natural splines it is possible to extrapolate to extreme values, without creating some unrealistic artefacts, like arbitrary and impossible swings, resulting for example from curve fitting methods based on higher degree polynomials. Natural cubic splines, which are constructed of piecewise third-order polynomials with continuity conditions expressed until second derivatives (Hastie and Tibshirani, 1990), are constrained to be linear outside the range of the data, and therefore provide a useful tool for extrapolation purposes. Note that requiring linearity outside the range of the data imposes additional smoothness constraints inside the range. Within the R package `mgcv` (Wood, 2006) recent developments of GAMs have been implemented, which have been applied in this study also utilizing the following command: `gam(y ~ s(x, bs="cr", k=10))`. In this `gam` function the smooth term of the observations x is specified by `s`. By setting the term `bs="cr"`, `k=10` the penalized cubic regression spline will be applied with the dimension `k` of the basis, used to represent the smooth term, and which sets the upper limit on the degrees of freedom.

4 Results

In order to evaluate the dependence of the NQT on the sample size and its effect on the predictive uncertainty based on the meta Gaussian distribution, two different cases have been analysed with respect to (a) extreme values (b) GAM based and (c) combined extreme values plus GAM based extrapolation. The two different cases are:

1. Small sample size (six years) at station Bohumin including no severe flood events

2. Sample at station Hofkirchen

   – covering eight years (i.e. the complete data set available, including severe events)

- after elimination of the first half of the data set (reduced time series excluding most of the severe events, which are concentrated in the first half of the observed data set)

5
- after elimination of the second half of the data set (most of the severe events remain included)

4.1 Station Bohumin – 6 yr

The available data set at station Bohumin has been very limited and the sample includes almost no severe events. Therefore the flood event in May 2010 is well suited for testing extrapolation methods in view of the NQT and deriving rules for application.
10 In the historical data set the observed maximum was about 600 $\mathrm{m^3\,s^{-1}}$, whereas in May 2010 the stream-flow values reached a maximum of approximately 1000 $\mathrm{m^3\,s^{-1}}$, which corresponds to a flood event with a return period of approximately twenty years (information provided by the Czech Hydrometeorological Institute, CHMI). Starting from the situation shown in Fig. 2a the historical data set has first been extended including
15 extreme values derived from the POT model. The non-linear transformation function is shown in Fig. 3a with the extended sample including extreme values shown as green dots. Applying to this "statistically" extended data set the NQT, the post-processor will result in the corrected forecast including the predictive uncertainty as shown in Fig. 5a. Comparing the predictive uncertainty estimated by the GAM based inter- and extrap-
20 olation of the non-linear back-transformation function of the NQT (Fig. 5b) with the enormous range of uncertainty in the forecast of the extreme value extended sample the drawback of this method becomes quite obvious. Although this huge uncertainty range corresponds quite well with the large uncertainty range of the extreme values estimated by the POT method (Fig. 4a), from the decision maker's point of view the worth
25 of the outcome of such an uncertainty band becomes questionable and will probably not be accepted from the end-users of flood forecast system. From the comparison of the three different approaches in Fig. 5 one can see that the GAM based uncertainty range is by far smaller than the ranges derived from methods including extreme values,

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

8, 9275–9297, 2011

Application of the NQT

K. Bogner et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

which could lead to an unrealistic impression of sharpness of the forecast system, i.e. not explaining all possible sources of uncertainty intrinsic to the flood forecasts. The last method of extrapolation is the combination of (a) extending the sample by extreme values estimated by the POT model and (b) GAM fit and extrapolation of this extended sample. In Fig. 5c the result of this method labelled "GAM + POT" is shown. The advantage of this method is the smooth transition between observed and extrapolated extreme values as opposed to the result of extending the sample size with extreme values only without applying GAM, which could result in sharp jumps in uncertainty as soon as the estimated values exceed the historical observed maximum value. In Fig. 5a this drastic discontinuity occurs at time-step zero (at forecast initiation), where the innermost quantile range (corresponding to the uncertainty range between 0.45 and 0.55 and shown by the darkest shaded area) covers a range of more than $500\,\mathrm{m}^3\,\mathrm{s}^{-1}$. Furthermore it can be seen how a slight difference in the stream-flow value in the normal space can cause a big difference in the real world after back-transformation, like the differences in the DWD deterministic forecast (based on weather forecasts provided by the German Weather Service) ranging from $1000\,\mathrm{m}^3\,\mathrm{s}^{-1}$ in Fig. 5a,b to $1300\,\mathrm{m}^3\,\mathrm{s}^{-1}$ in Fig. 5c. However in Fig. 5c it can be seen that the inner quantile ranges (0.35–0.65) cover all the observations in the forecast period (from leadtime 1 to 10) and the predicted median follows quite well the observations, whereas the first two methods result in under or over predictions from a leadtime of three days onwards.

4.2 Station Hofkirchen

The historical data sample at station Hofkirchen comprises eight years and includes several quite severe flood events such as those in May 1999, spring 2002 and August 2002, which correspond to flood events with return periods between 10 and 20 yr (information provided by the Bavarian Environmental Agency). Given this representative data set several tests have been conducted analysing the NQT performance and its dependence on sample size and on the extrapolation method. Since the majority of the flood events occurred in the first half of the sample (between 1998 and 2003) the

NQT and the post-processor have been applied on the total sample and on the split (halved) sample (Fig. 2b–d). Taking the total data set of eight years for calibrating the post-processor result in a forecast with a predictive uncertainty covering almost all the observations, although the first peak of this double-peaked event is overestimated and the second one, the more severe peak, is underestimated (Fig. 2b). Nonetheless the observed values of the second peak fall within the upper quantile range (0.05–0.95) at a leadtime of eight days, which is quite good for a medium range forecast. Fitting the post-processor to the first half of the sample size (Fig. 2c) leads to an increased uncertainty because of the greater variability of the sample. In comparison to this the calibration based on the second half sample result in to a forecast with smaller uncertainty bands because of the smaller variance of the sample, demonstrating the importance of sample size and explaining the upper limit of applicability of the NQT, when the observations exceed the historical sample (Fig. 2d).

In the case of the eight years data set, the artificial extension and/or the application of the GAM approach only resulted in modest changes in the forecast and consequently these results are not shown here. However taking only the first half of the sample with several severe events results in an extreme value distribution producing too heavy tails and therefore heavily overshooting predictive uncertainty ranges in the forecast example (Fig. 6a). The inclusion of extreme values fitted to the second half sample results in back-transformed forecast values which substantially exceed the observations because of the discontinuity between the historical data set and the estimated extreme values. Such discontinuities will result in unrealistic, abrupt changes as can be seen in Fig. 6b at leadtime three and can be circumvented by the application of the GAM in combination with the POT model, which will result in smooth forecasts (Fig. 6c).

5 Conclusions

In this study of the applicability of the NQT in flood forecasting systems different problems arising from the small sample sizes are discussed. The chosen forecast examples

at two different stream-flow gauging stations and for two different flood events demonstrate the problems of extending the historical data set by extreme values resulting from the fitted POT model. Because of the discontinuity between the observed historical sample and the estimated extreme values, sharp and unrealistic rises (or falls) in the hydrograph can occur after transforming the forecast from the Gaussian to the "real world" space. The analysed GAM for approximating the non-linear (back)transformation function could be an alternative, but the problem of possibly unrealistic small predictive uncertainty ranges has to be investigated in more detail with longer time-series and at different stations. However for these very limited data sets analysed the suggested way would be the combination of the extension of the small samples by extreme values and the inter- and extrapolation of this prolonged data set by the GAM, which results in smooth forecast hydrographs and not too optimistic under-dispersive predictive uncertainty ranges.

References

Ailliot, P., Thompson, C., and Thomson, P.: Mixed methods for fitting the GEV distribution, Water Resour. Res., 47, W05551, doi:10.1029/2010WR009417, 2011. 9281

Ashkar, F. and Tatsambon, C. N.: Revisiting some estimation methods for the generalized Pareto distribution, J. Hydrol., 346, 136–143, 2007. 9282

Bartholmes, J. C., Thielen, J., Ramos, M. H., and Gentilini, S.: The european flood alert system EFAS Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrol. Earth Syst. Sci., 13, 141–153, doi:10.5194/hess-13-141-2009, 2009. 9278

Bo, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine

basin climate for hydro-meteorological studies, Int. J. Climatol., 1655, 1643–1655, 2007. 9277

Bogner, K. and Pappenberger, F.: Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system, Water Resour. Res., 47, W07524, doi:10.1029/2010WR009137, 2011. 9276, 9278

Bogner, K., Cloke, H., Pappenberger, F., De Roo, A., and Thielen, J.: Improving the evaluation of hydrological multi-model forecast performance in the Upper Danube Catchment, Int. J. River Basin Manage., in press, 2011. 9282

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics, Springer, London, 2001. 9280

Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds, J. Roy. Stat. Soc. Ser B, 52, 393–442, 1990. 9282

Deutsch, C. and Journel, A.: GSLIB: Geostatistical Software Library and User's Guide, 2nd edn., Oxford University Press, New York, NY, 1997. 9276

Engeland, K., Hisdal, H., and Frigessi, A.: Practical extreme value modelling of hydrological floods and droughts: a case study, Extremes, 7, 5–30, doi:10.1007/s10687-004-4727-5, 2004. 9277

Finkenstädt, B. and Rootzén, H.: Extreme values in finance, telecommunications, and the environment, in: Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, 2004. 9280

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distributions of the largest or smallest member of a sample, Proc. Cambridge Philos. Soc., 24, 180–190, 1928. 9281

Gnedenko, B. V. and Kolmogorov, A. N.: Limit Theorems for Sums of Independent Random Variables, Addison-Wesley, Cambridge, 1949/1968, english translation in 1968, of original Russian edition from 1949. 9281

Goovaerts, P.: Geostatistics for Natural Resources Evaluation (Applied Geostatistics), Oxford University Press, New York, 1997. 9276

Gumbel, E. J.: Statistics of Extremes, 2nd edn., Columbia University Press, New York, 1958. 9281

Hastie, T. and Tibshirani, R.: Generalized additive models (with discussion), Stat. Sci., 1, 297–318, 1986. 9280

Hastie, T. and Tibshirani, R.: Generalized Additive Models, Chapman and Hall, London, 1990. 9281, 9283
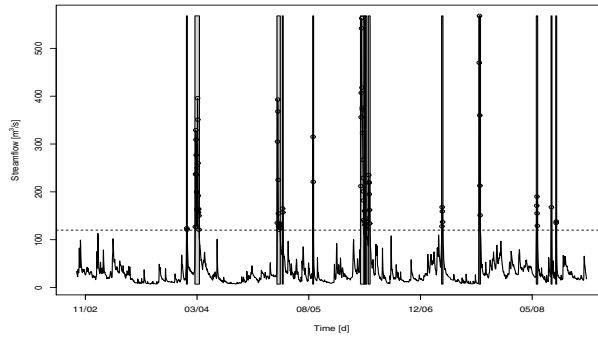
Hosking, J. and Wallis, J.: Parameter and quantile estimation for the generalized Pareto distribution, Technometrics, 29, 339–349, 1987. 9282

Kelly, K. and Krzysztofowicz, R.: A bivariate meta-Gaussian density for use in hydrology, Stochast. Hydrol. Hydraul., 11, 17–31, 1997. 9276

Krzysztofowicz, R.: Transformation and normalization of variates with specified distributions, J. Hydrol., 197, 286–292, 1997. 9277

Krzysztofowicz, R. and Herr, H. D.: Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model, J. Hydrol., 249, 46–68, 2001. 9276

Krzysztofowicz, R. and Kelly, K.: Hydrologic uncertainty processor for probabilistic river stage forecasting, Water Resour. Res., 36, 3265–3277, 2000. 9276, 9279

Krzysztofowicz, R. and Maranzano, C. J.: Hydrologic uncertainty processor for probabilistic stage transition forecasting, J. Hydrol., 293, 57–73, 2004. 9276

Laio, F., Porporato, A., Revelli, R., and Ridolfi, L.: A comparison of nonlinear flood forecasting methods, Water Resour. Res., 39, 1129, doi:10.1029/2002WR001551, 2003. 9281

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resour. Res., 45, W07416, doi:10.1029/2007WR006666, 2009. 9277

Lin, X. and Zhang, D.: Inference in generalized additive mixed modelsby using smoothing splines, J. Roy. Stat. Soc. Ser. B, 61, 381–400, doi:10.1111/1467-9868.00183, 1999. 9281

Luceño, A.: Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, Comput. Stat. Data Anal., 51, 904–917, 2006. 9282

MacKay, E., Challenor, P., and Bahaj, A.: A comparison of estimators for the generalised Pareto distribution, Ocean Eng., 38, 1338–1346, 2011. 9281

Mazas, F. and Hamm, L.: A multi-distribution approach to POT methods for determining extreme wave heights, Coast. Eng., 58, 385–394, 2011. 9281

McCullagh, P. and Nelder, J. A.: Generalized Linear Models, 2nd edn., Chapman & Hall, London, 1989. 9281

Moloney, N. and Davidsen, J.: Extreme bursts in the solar wind, Geophys. Res. Lett., 38, L14111, doi:10.1029/2011GL048245, 2011. 9281

Montanari, A.: Deseasonalisation of hydrological time series through the normal quantile transform, J. Hydrol., 313, 274–282, 2005. 9277

Moran, P.: Simulation and evaluation of complex water systems operations, Water Resour. Res., 6, 1737–1742, 1970. 9276

Nelder, J. A. and Mead, R.: A simplex method for function minimization, Comput. J., 7, 308–313, 1965. 9282

Pickands, J.: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, 1975. 9282

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/, ISBN:3-900051-07-0, 2011. 9276

Reggiani, P., Renner, M., Weerts, A., and van Gelder, P.: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, Water Resour. Res., 45, W02428, doi:10.1029/2007WR006758, 2009. 9276

Ribatet, M. A.: A User's Guide to the POT Package (Version 1.0), available at: http://cran.r-project.org/, 2006. 9281

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. 9277

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125–140, doi:10.5194/hess-13-125-2009, 2009. 9278

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Int. J. River Basin Manage., 6, 123–137, 2008. 9276

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: from dry events to heavy rainfalls, Water Resour. Res., 43, W07402, doi:10.1029/2006WR005308, 2007. 9277

Vrac, M. and Naveau, P.: Correction to "Stochastic downscaling of precipitation: From dry events to heavy rainfall", (Water Resour. Res., 43, (W07402) doi:10.1029/2006WR005308, 2007), Water Resour. Res., 44, W05702, doi:10.1029/2008WR007083, 2008. 9277

Wood, S.: Modelling and smoothing parameter estimation with multiple quadratic penalties, J. Roy. Stat. Soc. Ser. B, 62, 413–428, 2000. 9281

Wood, S.: Generalized Additive Models: An Introduction with R, Chapman and Hall/CRC, Boca Raton, 2006. 9281, 9283

Yu, P., Chen, S., and Chang, I.: Support vector regression for real-time flood stage forecasting, J. Hydrol., 328, 704–716, 2006. 9281

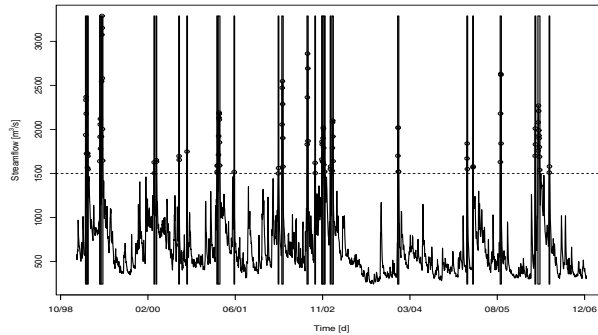Zhu, J.: A discussion on the extrapolation of hydrologic series, J. Hydrol., 96, 225–235, 1987. 9277

Application of the NQT

K. Bogner et al.

Table 1. River Gauging station information for Bohumin and Hofkirchen including stream-flow characteristic value MQ (mean discharge) and MHQ (average yearly mean discharge).

Station	River	Area (km^2)	Channel gradient $(m\,m^{-1})$	MQ $(m^3\,s^{-1})$	MHQ $(m^3\,s^{-1})$	Latitude $(°)$	Longitude $(°)$
Bohumin	Oder	4350	0.008	40	350	49.92	18.33
Hofkirchen	Danube	48 000	0.005	640	1870	48.68	13.11

(a) Bohumin

(b) Hofkirchen

Fig. 1. Observed stream-flow and number of clusters exceeding a threshold at station **(a)** Bohumin (Odra, CZ) and **(b)** Hofkirchen (Danube, DE) with 6, resp. 8 yr of daily observations.
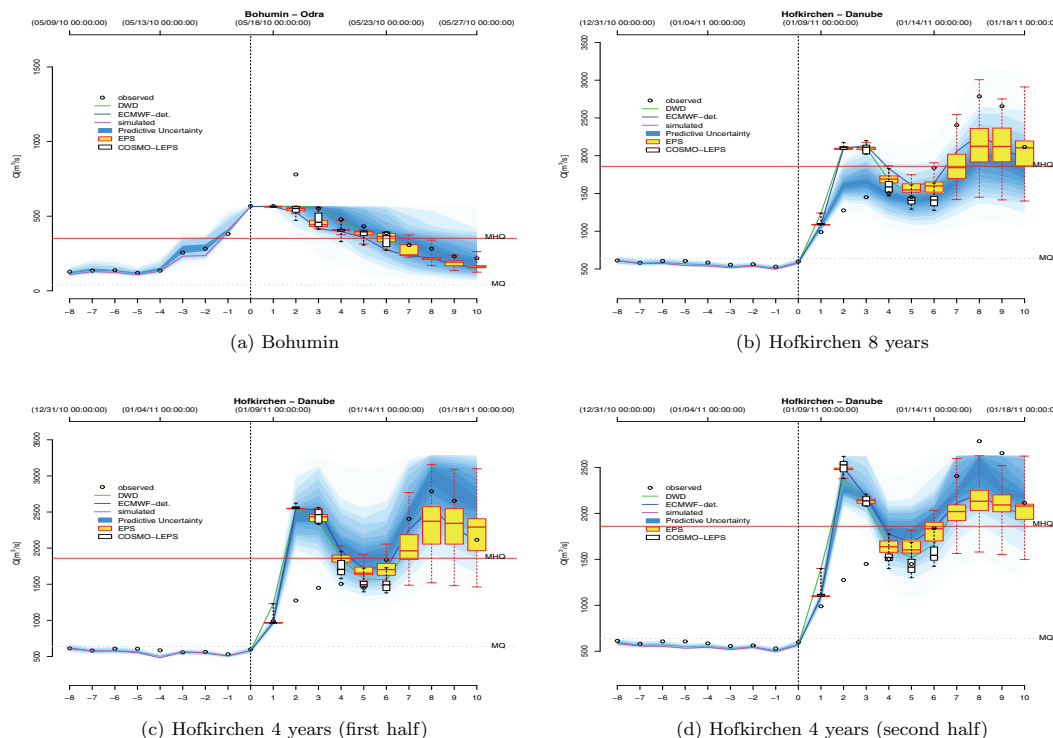
Fig. 2. Corrected forecast of a flood event in May 2010 at station Bohumin **(a)**, and in January 2011 at Hofkirchen **(b–d)** without extrapolation. At station Bohumin **(a)** and in the second reduced sample at station Hofkirchen **(d)** the upper limit of applicability is shown, when the maximum of the historical sample is exceeded during the forecast period.
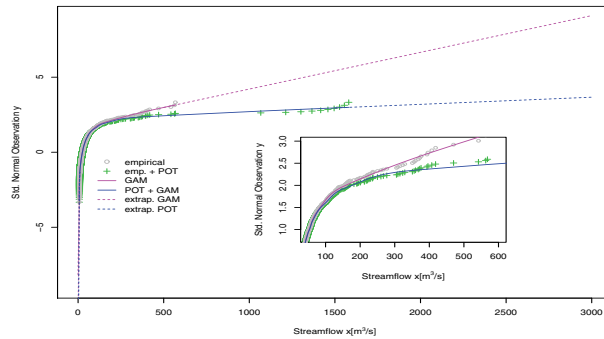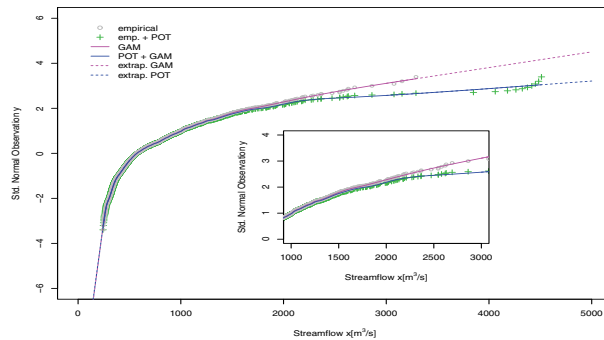
(a) Bohumin

(b) Hofkirchen 8 years

Fig. 3. Normalization of stream-flow values X at station Bohumin **(a)** and at Hofkirchen – 8 yr **(b)**. The range of observations with and without including extreme values (labelled as "emp. + POT", resp. "empirical") has been extrapolated applying GAMs.
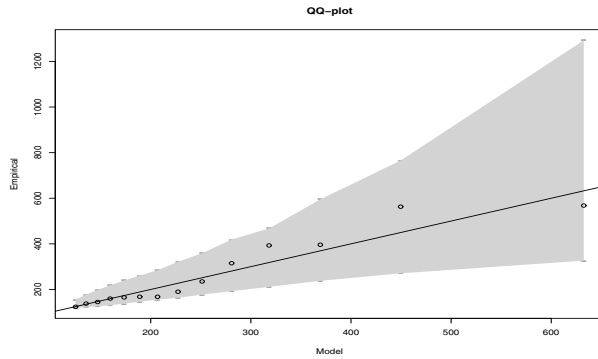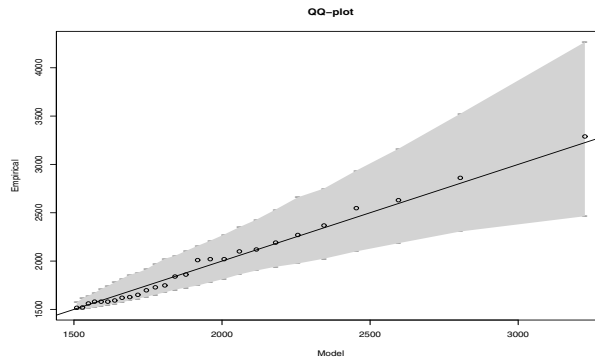
(a) Bohumin

(b) Hofkirchen 8 years

Fig. 4. Resulting QQ plot for the fitted extreme value distribution (POT Model) versus empirical extreme values including the 95 % confidence interval in gray for Bohumin **(a)** and Hofkirchen – 8 yr **(b)**.

(a) Bohumin - including extreme values

(b) Bohumin - GAM

(c) Bohumin - GAM + POT
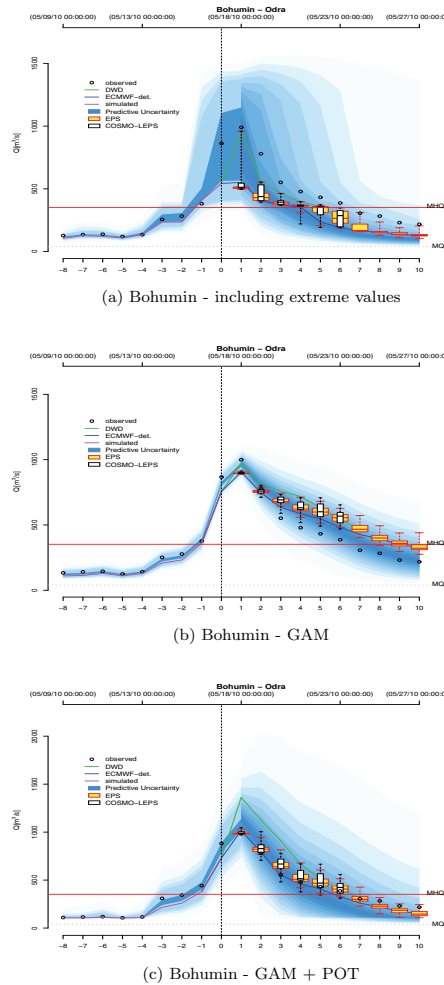
Fig. 5. Resulting forecast at station Bohumin applying three different methods of extrapolation.

(a) Hofkirchen 4 years (first half, POT)

(b) Hofkirchen 4 years (second half, POT)

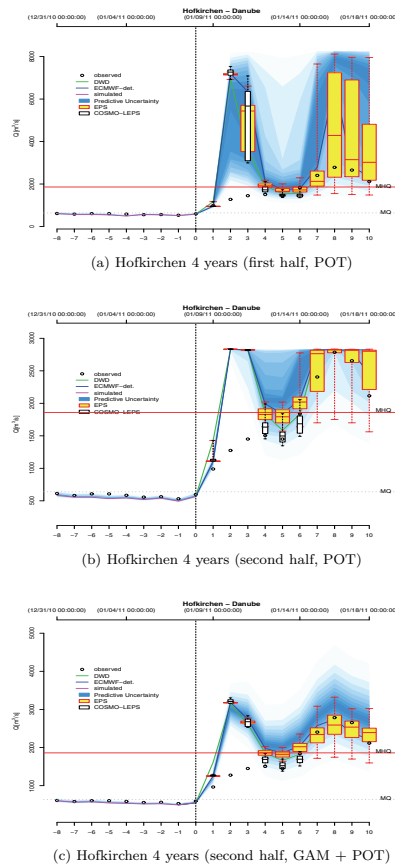(c) Hofkirchen 4 years (second half, GAM + POT)

Fig. 6. Corrected forecast at station Hofkirchen including extreme values for extrapolation based on **(a)** four years of the first half, including several severe events and **(b)** four years of the second half (less severe events) and **(c)** four years of the second half sample combining the GAM and the POT model for extrapolation.