

Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system

K. Bogner¹ and F. Pappenberger²

Received 23 January 2010; revised 3 May 2011; accepted 11 May 2011; published 13 July 2011.

[1] River discharge predictions often show errors that degrade the quality of forecasts. Three different methods of error correction are compared, namely, an autoregressive model with and without exogenous input (ARX and AR, respectively), and a method based on wavelet transforms. For the wavelet method, a Vector-Autoregressive model with exogenous input (VARX) is simultaneously fitted for the different levels of wavelet decomposition; after predicting the next time steps for each scale, a reconstruction formula is applied to transform the predictions in the wavelet domain back to the original time domain. The error correction methods are combined with the Hydrological Uncertainty Processor (HUP) in order to estimate the predictive conditional distribution. For three stations along the Danube catchment, and using output from the European Flood Alert System (EFAS), we demonstrate that the method based on wavelets outperforms simpler methods and uncorrected predictions with respect to mean absolute error, Nash-Sutcliffe efficiency coefficient (and its decomposed performance criteria), informativeness score, and in particular forecast reliability. The wavelet approach efficiently accounts for forecast errors with scale properties of unknown source and statistical structure.

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

1. Introduction

[2] Reliability, accuracy and timeliness are essential prerequisites for an efficient flood forecasting system. In addition, the translation of all available information (prior knowledge and model forecast) into reality, through the estimation of the predictive uncertainty, is necessary to allow decision makers to make the most effective decision under conditions of uncertainty. *Rougier* [2007] and *Todini* [2009] define predictive uncertainty as a subjective assessment of the probability of occurrence of a future (real) event, conditional upon all knowledge available up to the present and the information that can be acquired through a learning inferential process. The objective of this paper is to demonstrate possible advances regarding the efficiency of a hydrological forecasting system by reducing the prior uncertainty through analysis and correction of the error of the predictions and by providing the decision maker with estimates of the predictive uncertainty.

[3] A fundamental step to improve the reliability of the forecast is to analyze the prediction error in order to identify various sources of error, and to correct the error in order to minimize the error of the timing (timeliness), and the difference in the volume and magnitude of the peak between observed and simulated flood events (accuracy). Imperfec-

tions of hydrological predictions stem from uncertainties in the initial conditions, the boundary conditions, the forcing as well as model structure, and parameterization [e.g., see *Kachroo and Liang*, 1992; *Butts et al.*, 2004; *Cloke and Pappenberger*, 2009]. While it is important to improve and optimize a hydrological modeling chain by taking account of these imperfections through the calibration of the hydrological model (i.e., off-line calibration with historical data), operational forecasting also employs error correction, i.e., real-time compensation of the differences between the simulated and the most recent measured streamflow values.

[4] The various updating methods that are adopted can be classified into four groups, depending on the model variables modified during the updating process [*Georgakakos and Smith*, 1990; *World Meteorological Organization*, 1992]: (1) updating of input parameters, (2) updating of state variables, (3) updating of model parameters, and (4) updating of output variables. The last method is probably the most popular among hydrologists and has been successfully applied in flood forecasting because of its simple structure (e.g., autoregressive model) and its minimal computational costs (i.e., no time-consuming reruns of the hydrological forecast model are necessary), which are most relevant for real-time forecast systems.

[5] *O'Connell and Clarke* [1981], *Refsgaard* [1997] and *Xiong and O'Connor* [2002] report methods for model updating such as state variable updating based on an extended Kalman filter, methods based on neural networks and fuzzy logic or simple autoregressive and autoregressive threshold methods. *Xiong and O'Connor* [2002] state that efficiency of the simple standard autoregressive model is not improved by any of the more complicated methods that they have tested. The choice of any one method depends on what is

¹Land Management and Natural Hazards Unit, Institute for Environment and Sustainability, European Commission Joint Research Centre, Ispra, Italy.

²European Centre for Medium-Range Weather Forecasts, Reading, UK.

considered to be the main cause of discrepancy between observed and simulated streamflow values. The methods above ignore the possible scale dependence of the streamflow and error time series, i.e., the dependence upon the behavior of the process at the different temporal resolution under consideration. However, understanding the behavior of river flow processes and of derived quantities, such as the error time series, at different temporal (and spatial) scales is important in order to determine the predictability of river flow at different scales [Regonda *et al.*, 2004]. A highly suitable method of investigating the properties of time series at various resolutions (i.e., scales) is through the application of wavelet theory [Daubechies, 1992; Percival and Walden, 2000; Nason, 2008], which will be tested in this study.

[6] Error correction routines reduce the differences between observed and simulated time series significantly. The remaining modeling uncertainty (caused primarily by the underlying simplifications in the hydrological model, parameter uncertainty and errors in the initialization of the model runs such as measurement errors) can be evaluated in the context of a Hydrological Uncertainty Processor (HUP), which estimates the predictive uncertainty given observed streamflow measurements and (corrected) hydrological model predictions [Krzysztofowicz and Kelly, 2000]. In principal any postprocessing method which produces conditional probability distributions like the Model Output Statistic (MOS) method [Glahn and Lowry, 1972] and its modifications [Gneiting *et al.*, 2005; Wilks, 2009] could be applied. However, the advantage of the HUP based on Bayesian principles is that information not only from rather short historic pairs of observations and forecasts (predictions), but also from long climatic samples can be extracted and fused according to the laws of probability [Krzysztofowicz and Maranzano, 2006].

[7] The application of correction methods without analyzing the impact on the predictive uncertainty and vice versa (i.e., assessing the predictive uncertainty without updating the prior knowledge through real-time error correction) would result in an incomplete forecast system. It is essential to test a methodology in a full forecasting framework in order to assist decision makers to make the most efficient decision with the most reliable forecast while recognizing uncertainty [Krzysztofowicz, 1983, 2001a].

[8] For reasons of clarity, this paper shows the methodology for the error correction and the HUP only. The results of the application of the remaining parts of the Bayesian Forecast System (BFS), namely the *Input Uncertainty* resulting from the uncertainty of the meteorological (ensemble) forecast [Kelly and Krzysztofowicz, 2000; Herr and Krzysztofowicz, 2010] and the *Integrator* of the HUP and the *Input Uncertainty* [Krzysztofowicz, 2001b] are under preparation. The fundamental details about the BFS can be found in the work by Krzysztofowicz [1999a, 1999b].

[9] In this paper a novel approach to correct for errors in discharge is presented, that uses wavelet transformations and takes account of the scale dependences of errors. Its efficiency will be investigated in comparison with other simpler approaches and will be embedded into the BFS in order to estimate the total predictive uncertainty. At a first step the newly developed methodology of error correction in combination with the HUP will be tested within the framework of the European Flood Alert System [Thielen *et al.*, 2009], which uses meteorological forecasts to pre-

dict the probability of flooding in transnational river basins across Europe. First, in section 2, an overview of the applied methodologies (i.e., the theory of the wavelet transformation, the VARX and the HUP) will be discussed. This is followed in section 3 by a description of the hydrological system. In section 4 the evaluation methods are explained, followed by results and validation of the method in section 5. The conclusions are in section 6.

2. Applied Methodologies

2.1. Wavelet Transformation

[10] In recent years the development of wavelet theory has spawned applications in signal processing and time series analysis, and fast algorithms for integral transforms and image and function representation methods. A summary of the latest developments in the area of wavelet analysis and its applications can be found for example in the work by Bordeianu *et al.* [2009] and Qian *et al.* [2007]. Wavelet transformation provides a very useful and efficient decomposition tool for time series in terms of both time and frequency [see Mallat, 1989]. The objective of the approach presented here is to attempt to fit an error correction model to simulated discharge data and to update the corrected simulation in real time with observation data and weather forecasts. We acknowledge some similarities to the work of Chou and Wang [2004], who developed a rainfall runoff model directly in the wavelet domain. The approach in this paper is novel in that it adapts wavelet transformations as an error correction routine of a hydrological model using a different wavelet methodology, the Nondecimated Discrete Wavelet Transformation (NDWT), contrary to the Discrete Wavelet Transform (DWT) applied by Chou and Wang [2004] and Briggs and Levine [1997] also. Furthermore the emphasis of the latter work is on denoising and verifying meteorological field forecasts by decomposing spatial fields into scales representing different levels of details and applying multivariate scores. We note again that wavelets are particularly useful for analyzing the scale dependency of hydrological simulations and for the decomposition of the model errors into different temporal resolutions.

[11] In order to demonstrate the scale dependency of the time series the continuous wavelet transform (CWT) may be used. The CWT is a highly redundant mapping of a real-valued or complex-valued function $f(x)$ from the time domain to the so-called time-scale domain with CWT coefficients proportional to the variability of $f(x)$ at a given time and scale (for details see Carmona *et al.* [1998]). "Redundancy" means that a vast amount of repeated information is contained within this two-dimensional representation, depending on continuously varying translation and dilation parameters. Lane [2007] applied the CWT to rainfall runoff models and developed two objective functions based on wavelet power and wavelet phase in order to discriminate more effectively between different type of errors.

[12] Figure 1 illustrates the CWT of the time series of differences between observed and simulated discharges at the gauging station Hofkirchen (Upper Danube River, Germany) for the period 1 October 2001 to 30 September 2002. The magnitude of the CWT is plotted on a "scalogram," which is analogous to the spectrogram in time-frequency analysis, with time on the horizontal axis, scale

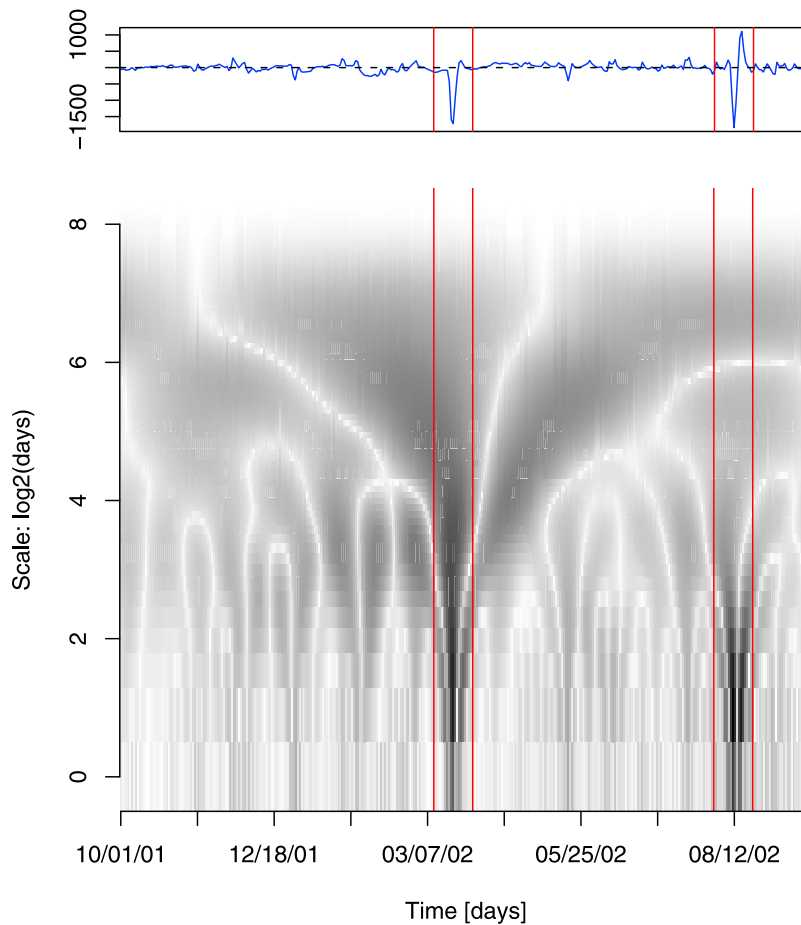


Figure 1. (top) Time series of error (observed minus simulated river discharge) at station Hofkirchen and (bottom) the time-scale plot of the corresponding continuous wavelet transformation. For display purposes the magnitude of the CWT coefficients is scaled by a factor of 0.5. The darkest regions, indicated by the boxes, highlight the regions of greatest variability having maximal CWT coefficients.

on the vertical axis, and amplitude given by grey scale. Two major error sources are observed in the scalogram. The largest is at springtime (in Figure 1 indicated with the first box at the end of March 2002), which is caused by long-lasting snowmelt processes. This is far larger than the amplitude of the error that appears in August 2002 (second box in Figure 1) which is caused by an overlapping of stratiform precipitation and convective rain fields of short duration. The effect of the springtime error lasts for more than two weeks (i.e., large wavelet coefficients up to and above the scale of 2^4 days), whereas the wavelet coefficients of the summer flood event disappear after a few days (indicated by the affected range of scales up to 2^3 days). For a more detailed discussion analysis on the relationship between the estimation of the wavelet coefficients and model diagnostics the reader is referred to *Briggs and Levine* [1997] and *Liu et al.* [2011].

[13] The wavelet-based method for the correction of the error between observed and modeled discharge series in the presented study is based on an alternative nondecimated wavelet transform, which is given by the à trous algorithm [Dutilleux, 1989] and has been applied for example by *Benaouda et al.* [2006] for forecasting purposes. The main objective of the proposed procedure will be the prediction of

future river discharge, and as such the most recent available observations are of critical importance.

[14] The observed and simulated time series are finite, of size say n , and the values at times $n, n-1, n-2, \dots$, are of greatest interest for predicting the next time steps. Any symmetric wavelet function is problematic for the handling of the boundary (or edge), and it will not be possible to use wavelet coefficients if these are calculated from unknown future data values. An asymmetric filter would be better for dealing with the edge of importance. Although one could hypothesize future data based on values in the immediate past, there is nevertheless discrepancy in fit in the succession of scales, which grows with scale as larger numbers of immediately past values are taken into account.

[15] In addition, for both symmetric and asymmetric functions a variant of the transform, that handles the edge problem, is needed. Usually this will be the mirror or periodic border handling. Although all of these methods work well in many applications, they are very problematic in forecast applications as they add artifacts in the most important part of the signal: its right border values. In order to address this boundary problem, *Bogner and Kalas* [2008] applied a redundant representation using dilations and translations of the autocorrelation functions of compactly

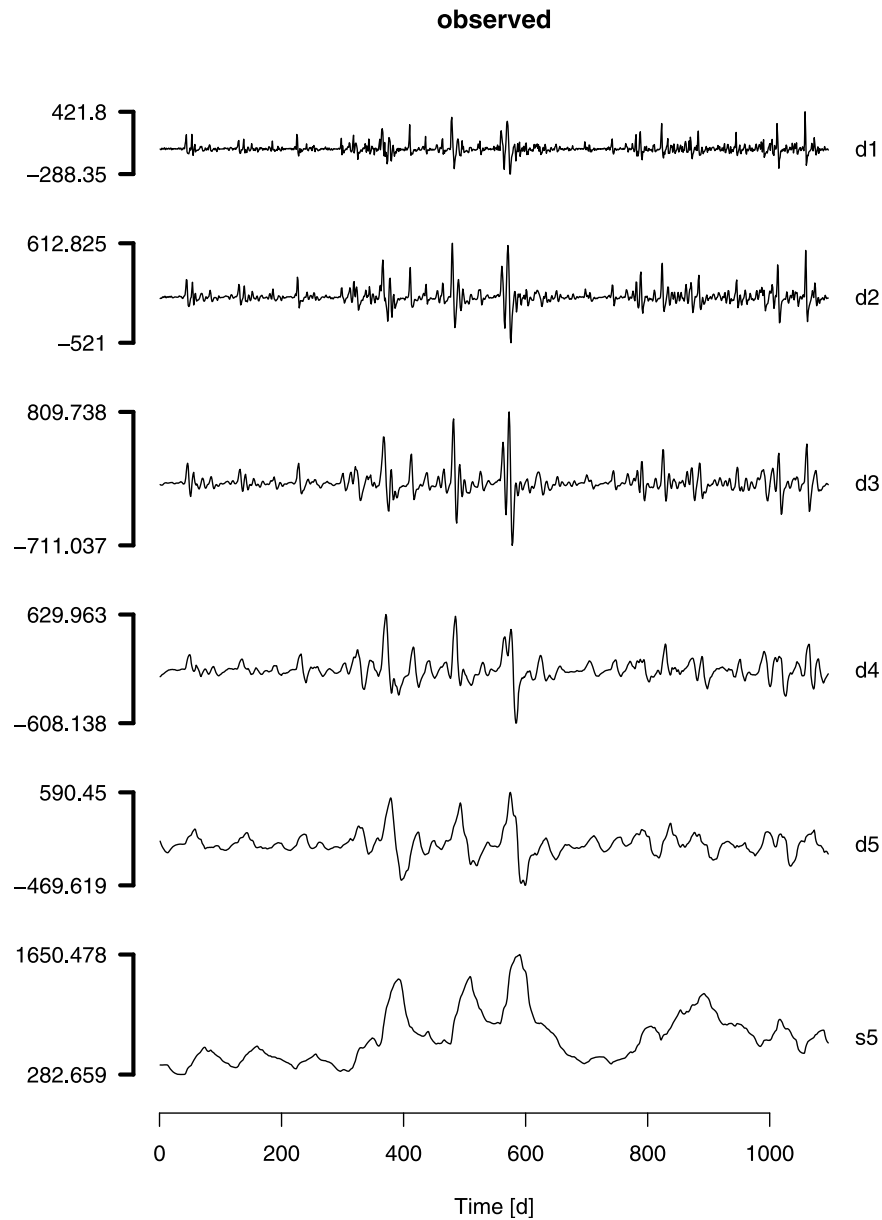


Figure 2. Nondecimated Haar wavelet transform with 5 levels of decomposition.

supported wavelets (the autocorrelation shell) instead of the wavelets per se [Saito and Beylkin, 1993]. However, in this paper the adaptive Haar à trous transform is tested, which is mathematically less complex and therefore straightforward to implement [Zheng et al., 1999; Renaud et al., 2003].

[16] The Haar wavelet transform may be considered as simply pairing up input values, storing the difference and passing the sum recursively, and pairing up the sums to provide the next scale (for a review see Stankovic and Falkowski [2003]). The Haar à trous wavelet transform is based on successive convolutions with the discrete low-pass filter $h(1/2, 1/2)$:

$$s_{i+1}(k) = \sum_{l=-\infty}^{\infty} h(l)s_i(k + 2^i l), \quad (1)$$

where the finest scale is the original signal $s_0(t) = y(t)$. From the sequence of smoothings of the signal the differences are used to calculate the wavelet coefficients d_i :

$$d_i(k) = s_{i-1}(k) - s_i(k), \quad (2)$$

which will capture the details of the signal.

[17] At any time point, k , no information after k is used for calculating the wavelet coefficients. The original time series y_t can be reconstructed as a linear combination of the wavelet and scaling coefficients

$$y_t = s_j(t) + \sum_{j=1}^J d_j(t). \quad (3)$$

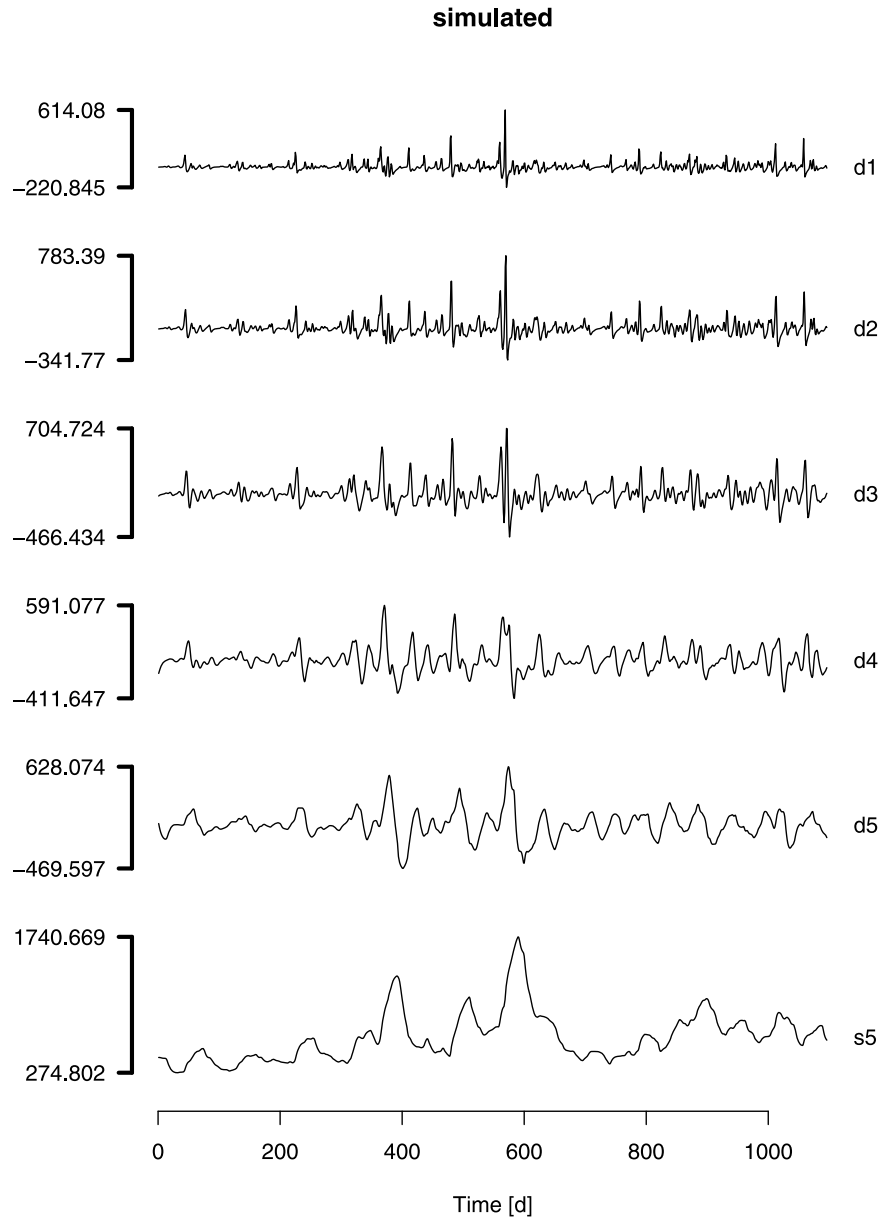


Figure 2. (continued)

[18] The decomposed observed and simulated series, as shown in Figure 2 for the Hofkirchen station (Upper Danube river, Germany), will be used to fit a times series model and to apply the error correction method. This is described in section 2.2.

2.2. Vector Autoregressive With Exogenous Inputs Model

[19] In this paper we apply a VARX model in order to predict the wavelet coefficients of the decomposed observed values, taking the series of decomposed simulations as exogenous input. The VARX model is the straightforward multivariate extension of the univariate ARX model allowing contemporaneously correlation of the variables with one another. In order to evaluate the performance of the VARX model updating procedure it will be compared with the ARX model combined with the most simple and widely

used Autoregressive (AR) updating procedure (details and derivations of the equations in this section are given in Appendix A). The VARX can be written in state-space form as

$$x_{t+1} = \phi x_t + \gamma u_t + w_t \quad t = 0, 1, \dots, n, \quad (4)$$

$$y_t = A_t x_t + \Gamma u_t + v_t \quad t = 1, \dots, n, \quad (5)$$

where, in the state equation, $x_0 \sim N(\mu_0, \Sigma_0)$, ϕ is $p \times p$, and γ is $p \times r$. In the observation equation, A_t is $q \times p$ and Γ is $q \times r$. Both w_t and v_t are white noise series, with $\text{var}(w_t) = Q$ and $\text{var}(v_t) = R$, and furthermore the state noise and observation noise are allowed to be correlated at time t (for derivations see *Shumway and Stoffer [2006]* and Appendix A).

[20] The conversion of the VARX model into state-space form allows the application of a reduction procedure based

on singular value decomposition [Mittnik, 1989; Aoki and Havenner, 1986]. This has the advantage that the resulting lower-dimensional model will be dynamically stable [Mittnik, 1990] and the method of least squares for fitting observed and simulated series of wavelet coefficients for a set of possible time lags can be applied. In addition the conversion to the state-space form allows the straight forward application of the recursive Kalman Filter state estimation algorithm given initial conditions $\mathbf{x}_{1|0}$ and filter error covariance $P_{1|0}$, for $t = 1, \dots, n$:

$$x_{t+1|t} = \phi x_{t|t-1} + \gamma u_t + K_t^* (y_t - A_t x_{t|t-1} - \Gamma u_t), \quad (6)$$

$$P_{t+1|t} = [\phi - K_t^* A_t] P_{t|t-1} [\phi - K_t^* A_t]' + Q + K_t^* R K_t^{*'} - S K_t^{*'} - K_t^* S', \quad (7)$$

where the new gain matrix is given by

$$K_t^* = [\phi P_{t|t-1} A_t' + S] [A_t P_{t|t-1} A_t' + R]^{-1}. \quad (8)$$

The filter update, given a new observation y_{t+1} and input u_{t+1} is given by

$$\begin{aligned} x_{t+1|t+1} &= x_{t+1|t} + P_{t+1|t} A_{t+1}' [A_{t+1} P_{t+1|t} A_{t+1}' + R]^{-1} \epsilon_{t+1}, \\ P_{t+1|t+1} &= P_{t+1|t} - P_{t+1|t} A_{t+1}' [A_{t+1} P_{t+1|t} A_{t+1}' + R]^{-1} \times A_{t+1} P_{t+1|t}. \end{aligned} \quad (9)$$

2.3. Hydrological Uncertainty Processor

[21] Section 2.2 introduced the VARX correction method, and later in this paper the quality of this method will be evaluated. However, it is important to undertake this evaluation in the context of the hydrological uncertainty. Ignoring this uncertainty would only give a partial, incomplete picture of the performances [Todini, 2009].

[22] Following the work of Krzysztofowicz [1999a], Krzysztofowicz and Kelly [2000], and Krzysztofowicz and Maranzano [2004] the HUP is developed based on the Bayesian formulation and a meta-Gaussian distribution family [Kelly and Krzysztofowicz, 1994, 1997]. Let y_n denote the observed discharge at a gauging station for lead time n ($n = 1, \dots, N$) prepared on day $n = 0$ and treated as a random variable Y_n , whereas r_n is viewed as a realization of R_n and is the output of a deterministic hydrological model denoting an estimate of Y_n , given a perfect weather forecast. Given observed meteorological input data, the first step is to convert the historical observed discharge values and the corresponding hydrological model predictions into Normal space using the quantiles associated with the order statistics [Krzysztofowicz, 1997; Kelly and Krzysztofowicz, 1997].

[23] The historical records of observed discharge values y and simulated discharge values r have been extended artificially by fitting the generalized Pareto distribution [Pickands, 1975] to the peaks-over-threshold (POT), and the estimated discharge values with low probabilities of occurrence are appended to the pairs of observation and simulation. This allows for the back transformation of extreme values which are not available in the limited sample size. An overview and further details about the POT modeling of hydrological extremes, and about the implications

of an assumption of the distribution, can be found in the work by Ashkar and Tatsambon [2007].

[24] The Normal Quantile Transform (NQT) of the variates Y_n and R_n is given by the composition of the inverse standard normal distribution Q and the empirical distribution Γ , resp. $\bar{\Lambda}$ of the variate:

$$W_n = Q^{-1}(\Gamma(Y_n)), \quad n = 0, 1, \dots, N, \quad (10)$$

$$X_n = Q^{-1}(\bar{\Lambda}_n(R_n)), \quad n = 1, \dots, N. \quad (11)$$

The next step is to formulate the *a priori* model, which will rest on the assumption that the NQ transformed discharge w_n follows a Markovian lag one process. Thus the process is governed by the normal linear equation

$$W_n = c W_{n-1} + \mathcal{E}, \quad (12)$$

where \mathcal{E} is normally distributed, $N(0, 1 - c^2)$ and the parameter c represents the Pearson's correlation coefficient. Also the likelihood function will rest on the assumption that the stochastic dependence between the transformed variates is governed by a simple normal linear equation $X_n = a_n W_n + d_n W_0 + b_n + \Theta_n$, where Θ_n is normally distributed with mean zero and variance σ_n^2 . Given the prior density and the likelihood function, the theory of conjugate families of distributions [De Groot, 1970] can be applied and the conditional expected density (likelihood function) can be derived:

$$\kappa_{Q_n}(x_n | w_0) = \frac{1}{(a_n^2 t_n^2 + \sigma_n^2)^{1/2}} q \left(\frac{x_n - (a_n c^n + d_n) w_0 - b_n}{(a_n^2 t_n^2 + \sigma_n^2)^{1/2}} \right), \quad (13)$$

where $t_n^2 = 1 - c^{2n}$ and q denote the standard normal density and subscript Q denotes a density in the space of transformed variates [Krzysztofowicz and Kelly, 2000]. According to the Bayesian theorem the posterior density can be derived

$$\phi_{Q_n}(w_n | x_n, w_0) = \frac{1}{T_n} q \left(\frac{w_n - A_n x_n - D_n w_0 - B_n}{T_n} \right), \quad (14)$$

where

$$A_n = \frac{a_n t_n^2}{a_n^2 t_n^2 + \sigma_n^2}, \quad B_n = \frac{-a_n b_n t_n^2}{a_n^2 t_n^2 + \sigma_n^2}, \quad (15)$$

$$D_n = \frac{c^2 \sigma_n^2 - a_n d_n t_n^2}{a_n^2 t_n^2 + \sigma_n^2}, \quad T_n^2 = \frac{t_n^2 \sigma_n^2}{a_n^2 t_n^2 + \sigma_n^2}. \quad (16)$$

The application of the HUP for operational flood forecasting purposes has the advantage that the fitting of the HUP to historical data can be calculated off-line and only a small set of estimated parameters (a_n , d_n , b_n and σ_n^2) will have to be stored. A step-by-step guide of the practical application of the HUP can be found in the work by Todini [2008].

3. Domain and Model Description

[25] The methodology will be tested within the context of the European Flood Alert System (EFAS). The aim of EFAS

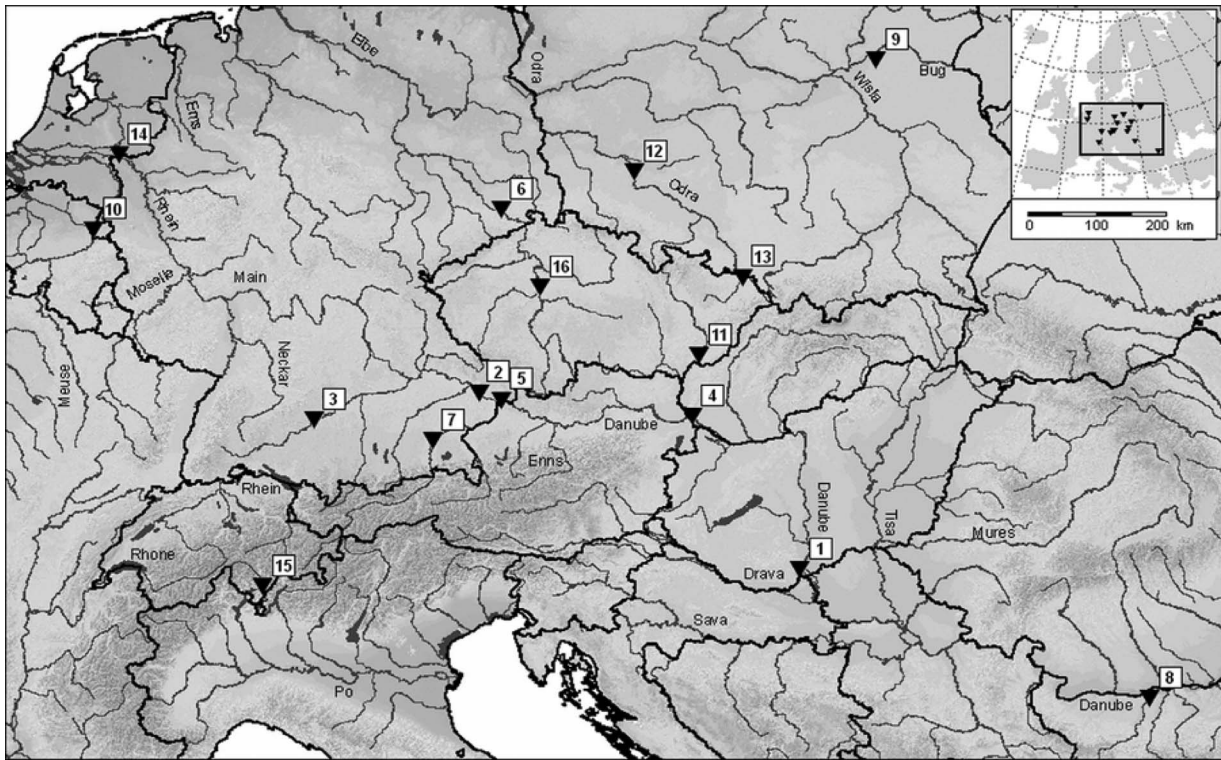


Figure 3. Sixteen stations running operational realtime postprocessing and updating in EFAS.

is to increase the preparedness for floods in transnational European river basins. A detailed setup of the system is described by *Thielen et al.* [2009] and *Bartholmes et al.* [2009]. EFAS uses the hydrological model LISFLOOD, which is a distributed, hydrological rainfall runoff model. It is a hybrid between a conceptual and a physical rainfall runoff model designed specifically to simulate the hydrological processes that occur in large catchments [*Van Der Knijff et al.*, 2010]. The model has been extensively tested and calibrated for various catchments across the globe [*Mo et al.*, 2006; *Feyen et al.*, 2007; *He et al.*, 2009; *Thiemig et al.*, 2010].

[26] In this study the operational setup of EFAS is used which covers Europe on a 5 km grid and simulates discharges on 6 hourly time steps. The model parameters for the simulation of the soil and land use processes are derived from European databases. Snowmelt is simulated using a degree day approach, with a correction factor for higher snowmelt rates during rain events. The model parameters that control snowmelt rates, infiltration, overland and river flow, as well as residence time in the soil and subsurface reservoirs, were calibrated with the SCE-UA algorithm [*Duan and Sorooshian*, 1992], replacing the original random sampling with a Latin Hypercube sampling scheme to generate the initial population. Following recommendations by *Feyen et al.* [2008], calibration was done in a semidistributed way, dividing each catchment into subcatchments (in total 231), based on available station data.

[27] The evaluation of the updating methodology will concentrate on three selected stations along the Upper Danube river (see Figure 3 and Table 1), which were the first ones to implement the postprocessing procedure. The

number of stations is steadily increasing up to 16 stations at present, where the updating is performed operationally twice a day. The streamflow at the three selected stations is dominated by convective precipitation in summer and snow melting in spring, and are representative for Central Europe, showing hydrological characteristics ranging from nival (like the Inn river joining the Danube river upstream Achleiten) to pluvial (like the Danube river upstream Neu Ulm).

[28] In Table 1 some hydrologically relevant numbers are summarized, where MQ stands for the mean annual daily discharge and MHQ refers to the mean annual maximum discharge. The MQ and MHQ values have been provided by local water authorities (as for the testing stations) or have been calculated based on available historical series of observed discharge data. Since EFAS incorporates station data from all over Europe, many different data formats and different time periods of historical data are available. Therefore the calibration period of the VARX is different for the selected station Achleiten (Austria) and for the two stations in Germany (Neu Ulm and Hofkirchen), but they all span at least a period of 3 years including severe flood events.

4. Evaluation Criteria

4.1. Deterministic Evaluation

[29] The quality of point prediction models, such as the deterministic output of a hydrological model, has been assessed with the mean absolute error (MAE) and the Nash-Sutcliffe (N-S) coefficient [*Nash and Sutcliffe*, 1970]. In order to measure the skill of the analyzed methods the

Table 1. Gauging Stations^a

ID	Station	Tributary/River	Area (km ²)	MQ (m ³ /s)	MHQ (m ³ /s)	Channel Gradient (m/m)
1	Mohacs	Danube	210,100	2,470	5,100	0.005
2	<i>Hofkirchen</i>	<i>Danube</i>	<i>48,000</i>	<i>640</i>	<i>1,870</i>	<i>0.005</i>
3	<i>Neu Ulm</i>	<i>Danube</i>	<i>8,000</i>	<i>125</i>	<i>588</i>	<i>0.009</i>
4	Bratislava	Danube	130,000	2,100	5,900	0.010
5	<i>Achleiten</i>	<i>Danube</i>	<i>72,500</i>	<i>1,430</i>	<i>4,120</i>	<i>0.015</i>
6	Dresden	Elbe	332,500	320	1,400	0.008
7	Wasserburg	Inn/Danube	11,600	360	1,450	0.034
8	Orahovica	Iskar/Danube	8,000	460	1,160	0.014
9	Lochow	Liwiec/Vistula	2,800	10	60	0.002
10	Borgharen	Maas	20,000	230	1,500	0.006
11	Straznice	Morava/Danube	10,700	70	380	0.008
12	Scinawa	Oder	30,000	180	790	0.004
13	Bohumin	Oder	4,350	40	350	0.008
14	Lobith	Rhine	160,000	2,300	6,600	0.010
15	Bellinzona	Ticino/Po	1,575	65	510	0.054
16	Prague	Vltava/Elbe	26,000	140	840	0.007

^aThe selected stations for demonstrating the quality of the updating performance are in italic font.

improvements in the MAE of the corrected predictions over the MAE of the uncorrected predictions will be calculated. Additionally the N-S coefficient was further decomposed into three distinctive components representing the correlation r , bias β and α as a measure of relative variability in the observed values [Gupta et al., 2009]:

$$N-S = 2\alpha r - \alpha^2 - \beta^2, \quad (17)$$

where α is given by the ratio of the simulated and observed standard deviation and $b = (\mu_s - \mu_o)/\sigma_o$, with the simulated and observed mean discharge μ_s and μ_o .

4.2. Evaluating the Assumptions of HUP and Predictive Capability

[30] The first step in the evaluation of the HUP is to check the meta-Gaussian model dependence structure of the prior density, which requires examination of the assumption of *linearity*, *homoscedasticity* and *normality*. The three items are judged graphically [Krzyzstofowicz and Kelly, 2000] by scatterplots of the transformed sample points (w_0 , w_1) to assess linearity, and of the regression residuals ϵ versus values of $E(W_1|W_0 = w_0)$ to assess homoscedasticity. The last item of normality can be checked by plotting the empirical distribution of \mathcal{E} . Some conclusions about the predictive capability of the hydrological model [Krzyzstofowicz and Kelly, 2000] can be drawn from the estimated HUP parameters, a_n , d_n , b_n and σ_n^2 .

[31] The slope a measures how much information is in the output (or the “signal” carried by the output), while σ^2 measures the “noise” of the output. The “signal to noise” ratio a^2/σ^2 could be interpreted as a measure of hydrological uncertainty, with $a^2/\sigma^2 \rightarrow \infty$ implying no uncertainty and $a^2/\sigma^2 \rightarrow 0$ indicating infinite uncertainty. A decision-theoretic measure of informativeness of the output has been derived [Krzyzstofowicz and Evans, 2008], namely the informativeness score (IS), also referred to as the Bayesian correlation score in the original publication [Krzyzstofowicz, 1992]:

$$IS = \left[\left(\frac{a}{\sigma} \right)^{-2} + 1 \right]^{-\frac{1}{2}}. \quad (18)$$

The advantage of IS in comparison to the score from the signal to noise ratio is that the IS ranges in value from 0 (= uninformative predictor) to 1 (= perfect predictor).

4.3. Probabilistic Evaluation

[32] The joint distribution of the observations and the forecasts contains all the information necessary for evaluating the quality of the forecast. In the work by Murphy and Winkler [1987, 1992] this general diagnostic framework for (probabilistic) forecast verification is described in detail, which is based on the probability distribution of forecasts and observations and on the conditional and marginal distributions associated with factorizations of the joint distribution. There are two possible ways of factorization, namely, (1) into conditional distributions of the observations given the forecasts and the marginal distribution of the forecasts and (2) into conditional distributions of the forecasts given the observations and the marginal distribution of the observations. The second type is called the likelihood base rate factorization and indicates the possible usefulness of a forecast in predicting the observation in comparison to predictions based on the sample climatology without forecasts. This factorization can provide additional insights into the forecast quality, however for streamflow series the base rate, reflecting the long-term climatology, is difficult to estimate, because of possible spatial and temporal peculiarities and inherent nonlinearities. Therefore the first factorization, which is called calibration-refinement factorization, will be applied. In this factorization the conditional distribution indicates how often different observations occurred given a particular forecast and relates to the calibration or *reliability* of a forecast. The predictive or marginal distribution of the forecasts relates to the refinement or *sharpness* of the forecasts [Murphy et al., 1989].

[33] Various measures of the predictive performance and the quality of probabilistic forecasts address reliability and sharpness simultaneously, such as the Continuous Ranked Probability Score (CRPS), which is equivalent to the Brier Score integrated over all possible threshold values for the continuous predictand [Hersbach, 2000]. Gneiting and Raftery [2007] show that CRPS can be interpreted as a generalized version of the MAE, which will be used as a measure of accuracy here, where accuracy is defined as the

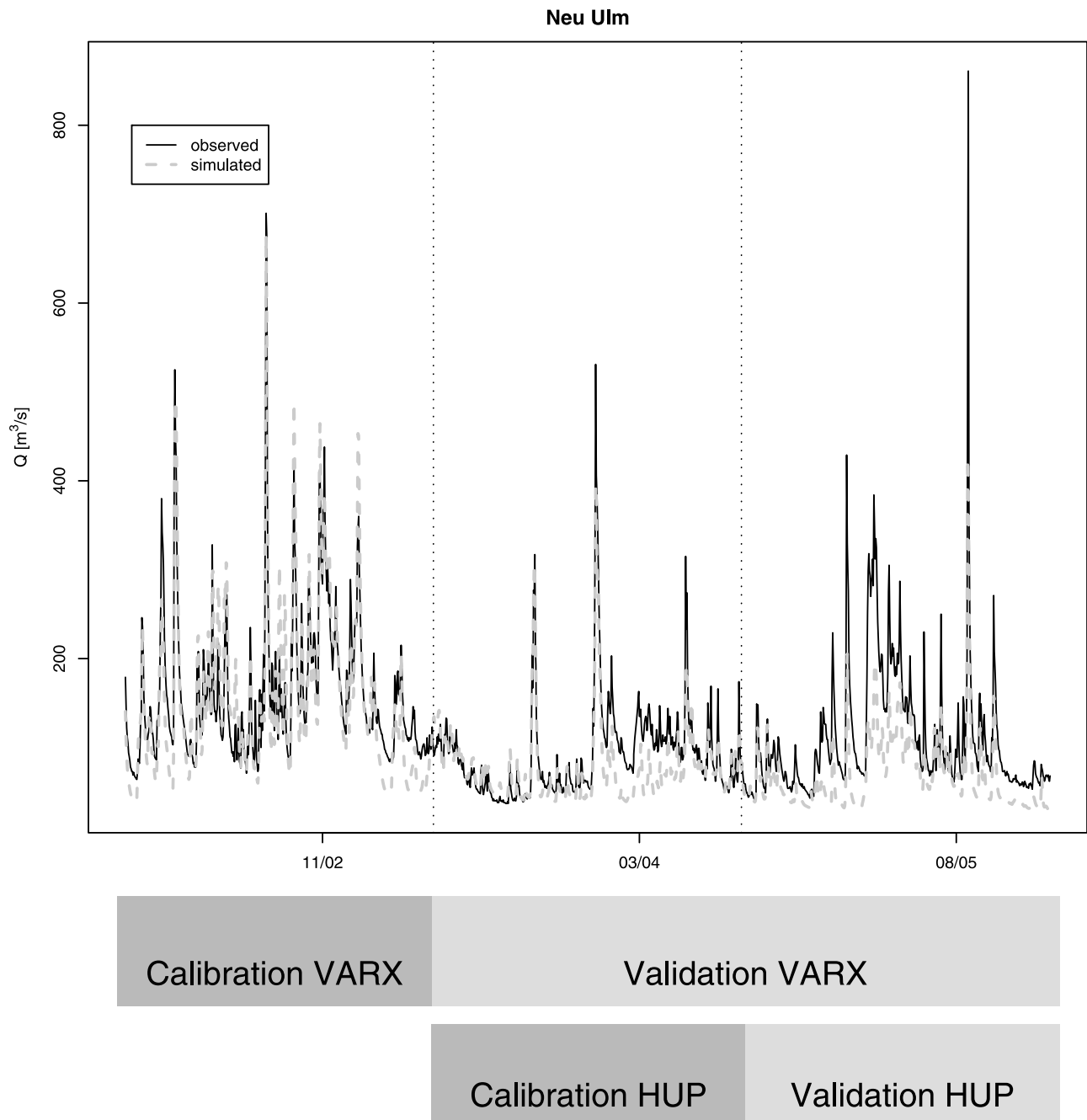


Figure 4. Observed and simulated river discharge at three Danube stations (Germany, Austria).

average degree of correspondence between individual forecasts and observation [Murphy and Daan, 1985; Murphy, 1988]. In this paper we use the Probability Integral Transforms or PIT (for details see Rosenblatt [1952]) of the forecasts in order to test for departures from the dual hypotheses of independence and uniformity, with departures from uniformity tested using the Kolmogorov-Smirnov (K-S) statistic [Conover, 1999; Noceti et al., 2003]. The rank test is a nonparametric test, used to test the null hypothesis H_0 that the input are values of independent and identically distributed random variables [see, e.g., Brockwell and Davis, 2002]. The PIT is defined as the value that the

predictive cumulative distribution function (CDF) attains at the observation. If F_t is the predictive distribution and x_t materializes, the transform is defined as $z_t = F_t(x_t)$. For a correct probability forecast the z_t values are mutually independent and identically $U(0, 1)$ distributed [Dawid, 1984; Berkowitz, 2001]. Diebold et al. [1998] suggested plotting a histogram of the resulting PIT values and comparing this to a perfect $U(0, 1)$ distribution as a method of discerning the degree of reliability of a model.

[34] Instead of plotting histograms the approach of Laio and Tamea [2007] will be used for producing predictive QQ plots, which does not require a subjective binning of the

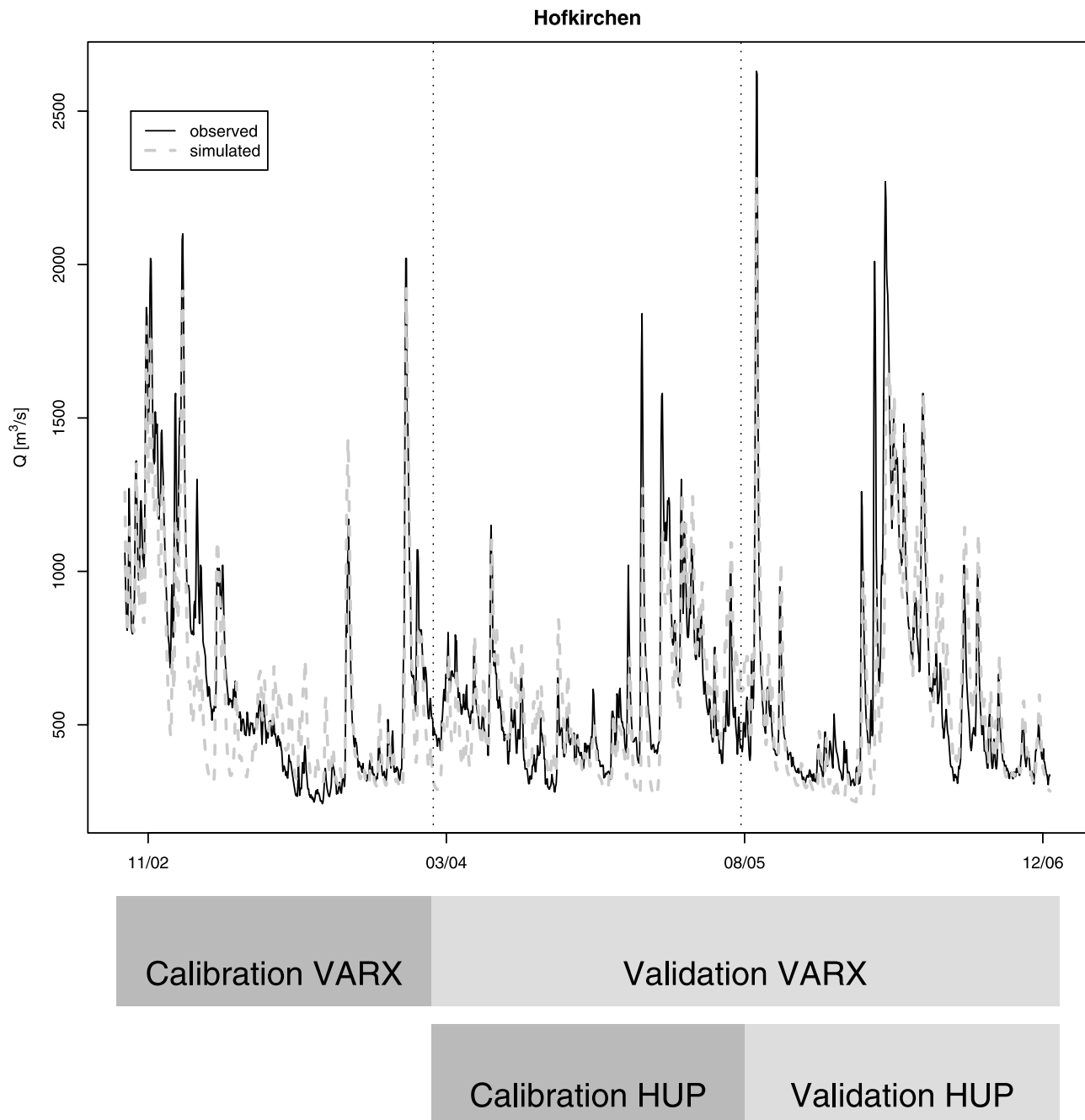


Figure 4. (continued)

data. Departures from the 1:1 line indicate that the observed values lie outside the predicted range, implying that the predictive uncertainty is underestimated or overestimated [Thyer *et al.*, 2009]. In order to evaluate the degree of departure the maximum absolute distance D_{\max} between the empirical distribution of the z_i values and the $U(0, 1)$ distribution can be calculated and will be used as a measure of reliability similar to the approach of Renard *et al.* [2010].

5. Results and Discussion

[35] For each station the series of paired observations and simulations has been divided into three parts (see Figure 4),

where the first part is used for the calibration of the VARX model in the wavelet domain and the second and third parts are used for the evaluation of the VARX model. The second part is also used for fitting the HUP and the third part is used for the evaluation of the predictive uncertainty. Therefore the following evaluations are separated according to this division.

5.1. Evaluation of the VARX

[36] In Table 2 the Nash-Sutcliffe (N-S) coefficients and the MAE for the uncorrected simulations and for the VARX, ARX and AR updating model for the three stations are shown for the verification period only, and demonstrate the

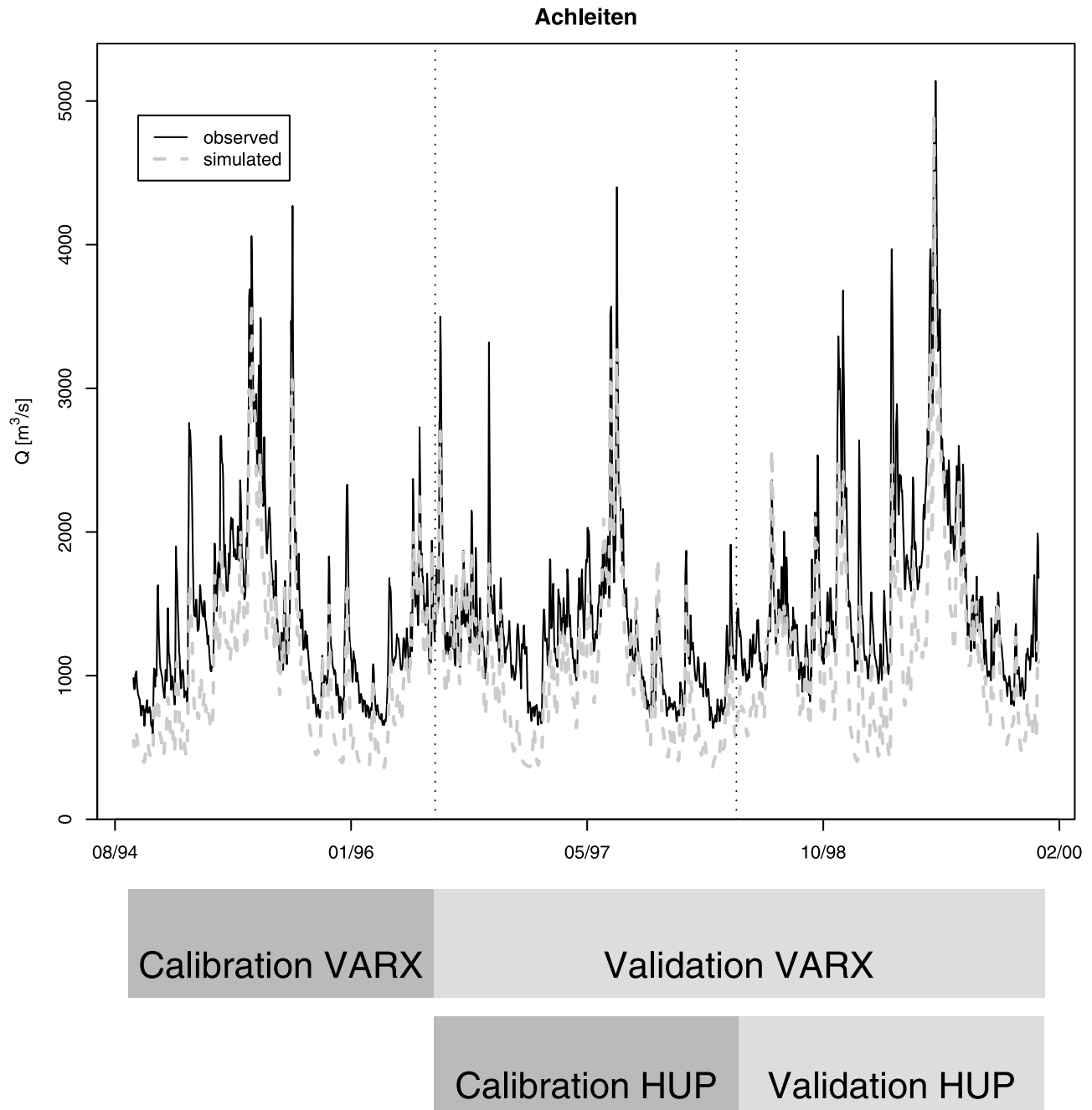


Figure 4. (continued)

improvements due to the wavelet transformation. In order to show the improved prediction performance more clear, a Prediction Skill Score (PSS) is given also, and is defined as the improvement in accuracy of the corrected predictions over the uncorrected predictions in analogy to the mean square error skill score of *Murphy* [1988]. Furthermore an example of the prediction performance at lead time three conditional on the streamflow magnitude is shown in Figure 5 demonstrating the superior performance of the VARX at almost all ranges of magnitude. A nonparametric resampling approach based on the *pairs bootstrapping* method [*Freedman*, 1981] is used in order to construct 95% confidence intervals (see A. Canty and B. D. Ripley, boot:

Bootstrap R (S-Plus) functions, R package version 1.3-2, 2011, and *Davison and Hinkley* [1997] for details). In Figure 5 the adjusted bootstrap percentile intervals [*DiCiccio and Efron*, 1996] are shown indicating the direct relationship of uncertainty and discharge magnitude.

[37] The MAE, N-S efficiency and its decomposed coefficients for the calibration of the updating procedure are shown in Table 3. Apart from the MAE, which indicates clearly the improvements achieved by the VARX model, the N-S coefficients and its decompositions are hard to distinguish for the period of calibrating the AR, ARX and VARX parameters. Regarding the validation period the VARX model is superior to the ARX and AR model for almost all lead times.

Table 2. Nash-Sutcliffe Efficiency Coefficient, Mean Absolute Error, and Prediction Skill Score (PSS) Taking the Mean Absolute Error of the Uncorrected Predictions as Reference Method^a

Measure	Method	Lead Time (days)									
		1	2	3	4	5	6	7	8	9	10
<i>Neu Ulm</i>											
N-S	uncorrected	0.44	0.44	0.44	0.44	0.43	0.43	0.43	0.43	0.42	0.42
	VARX	<i>0.93</i>	<i>0.87</i>	<i>0.82</i>	<i>0.79</i>	<i>0.77</i>	<i>0.75</i>	<i>0.73</i>	<i>0.71</i>	<i>0.69</i>	<i>0.67</i>
	ARX	<i>0.93</i>	<i>0.87</i>	<i>0.82</i>	<i>0.78</i>	<i>0.74</i>	<i>0.71</i>	<i>0.68</i>	<i>0.65</i>	<i>0.62</i>	<i>0.60</i>
	AR	<i>0.93</i>	<i>0.86</i>	<i>0.79</i>	<i>0.74</i>	<i>0.70</i>	<i>0.66</i>	<i>0.62</i>	<i>0.58</i>	<i>0.55</i>	<i>0.52</i>
MAE	uncorrected	32.8	32.8	32.8	32.79	32.8	32.8	32.8	32.8	32.8	32.8
	VARX	10.8	<i>14.2</i>	<i>17.1</i>	<i>18.7</i>	<i>19.8</i>	<i>20.6</i>	<i>21.2</i>	<i>21.8</i>	<i>22.3</i>	<i>22.6</i>
	ARX	<i>10.7</i>	14.3	17.6	19.5	20.8	22.0	23.0	23.7	24.5	25.0
	AR	11.6	15.1	18.3	20.1	21.4	22.6	23.6	24.4	25.1	25.6
PSS	VARX	<i>0.67</i>	<i>0.57</i>	<i>0.48</i>	<i>0.43</i>	<i>0.40</i>	<i>0.37</i>	<i>0.35</i>	<i>0.34</i>	<i>0.32</i>	<i>0.31</i>
	ARX	<i>0.67</i>	<i>0.56</i>	<i>0.46</i>	<i>0.41</i>	<i>0.37</i>	<i>0.33</i>	<i>0.30</i>	<i>0.28</i>	<i>0.26</i>	<i>0.24</i>
	AR	<i>0.65</i>	<i>0.54</i>	<i>0.44</i>	<i>0.39</i>	<i>0.35</i>	<i>0.31</i>	<i>0.28</i>	<i>0.26</i>	<i>0.24</i>	<i>0.22</i>
<i>Hofkirchen</i>											
N-S	uncorrected	0.46	0.46	0.46	0.47	0.47	0.47	0.48	0.48	0.49	0.49
	VARX	<i>0.97</i>	<i>0.94</i>	<i>0.91</i>	<i>0.89</i>	<i>0.86</i>	<i>0.85</i>	<i>0.83</i>	<i>0.83</i>	<i>0.82</i>	<i>0.82</i>
	ARX	<i>0.97</i>	<i>0.93</i>	<i>0.87</i>	<i>0.82</i>	<i>0.79</i>	<i>0.78</i>	<i>0.77</i>	<i>0.77</i>	<i>0.76</i>	<i>0.76</i>
	AR	<i>0.94</i>	<i>0.87</i>	<i>0.80</i>	<i>0.75</i>	<i>0.73</i>	<i>0.72</i>	<i>0.71</i>	<i>0.71</i>	<i>0.70</i>	<i>0.70</i>
MAE	uncorrected	118.7	118.6	118.5	118.4	118.3	118.1	118.0	117.8	117.7	117.6
	VARX	32.5	<i>40.5</i>	<i>52.2</i>	<i>62.2</i>	<i>69.8</i>	<i>75.3</i>	<i>78.7</i>	<i>80.7</i>	<i>81.1</i>	<i>81.4</i>
	ARX	33.8	49.2	66.0	78.3	86.6	89.9	92.6	94.0	94.8	96.2
	AR	50.3	69.1	85.5	93.2	96.9	97.6	97.8	95.7	94.4	93.1
PSS	VARX	<i>0.73</i>	<i>0.66</i>	<i>0.56</i>	<i>0.47</i>	<i>0.41</i>	<i>0.36</i>	<i>0.33</i>	<i>0.32</i>	<i>0.31</i>	<i>0.31</i>
	ARX	<i>0.71</i>	<i>0.58</i>	<i>0.44</i>	<i>0.34</i>	<i>0.27</i>	<i>0.24</i>	<i>0.21</i>	<i>0.20</i>	<i>0.19</i>	<i>0.18</i>
	AR	<i>0.58</i>	<i>0.42</i>	<i>0.28</i>	<i>0.21</i>	<i>0.18</i>	<i>0.17</i>	<i>0.17</i>	<i>0.19</i>	<i>0.20</i>	<i>0.21</i>
<i>Achleiten</i>											
N-S	uncorrected	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
	VARX	<i>0.96</i>	<i>0.91</i>	<i>0.88</i>	<i>0.85</i>	<i>0.84</i>	<i>0.83</i>	<i>0.82</i>	<i>0.81</i>	<i>0.80</i>	<i>0.79</i>
	ARX	<i>0.97</i>	<i>0.91</i>	<i>0.86</i>	<i>0.82</i>	<i>0.80</i>	<i>0.78</i>	<i>0.77</i>	<i>0.76</i>	<i>0.75</i>	<i>0.74</i>
	AR	<i>0.95</i>	<i>0.88</i>	<i>0.83</i>	<i>0.80</i>	<i>0.79</i>	<i>0.77</i>	<i>0.77</i>	<i>0.76</i>	<i>0.75</i>	<i>0.73</i>
MAE	uncorrected	364.5	365.3	365.8	366.3	366.6	366.9	367.3	368.0	368.4	368.5
	VARX	86.6	<i>123.1</i>	<i>144.3</i>	<i>155.6</i>	<i>162.4</i>	<i>169.5</i>	<i>173.4</i>	<i>174.9</i>	<i>178.2</i>	<i>181.1</i>
	ARX	78.4	124.5	152.7	167.1	177.3	185.9	190.3	193.9	197.3	199.6
	AR	92.8	144.3	170.3	183.5	190.7	195.5	196.8	197.0	200.5	204.0
PSS	VARX	<i>0.76</i>	<i>0.66</i>	<i>0.61</i>	<i>0.58</i>	<i>0.56</i>	<i>0.54</i>	<i>0.53</i>	<i>0.52</i>	<i>0.52</i>	<i>0.51</i>
	ARX	<i>0.78</i>	<i>0.66</i>	<i>0.58</i>	<i>0.54</i>	<i>0.52</i>	<i>0.49</i>	<i>0.48</i>	<i>0.47</i>	<i>0.46</i>	<i>0.46</i>
	AR	<i>0.75</i>	<i>0.60</i>	<i>0.53</i>	<i>0.50</i>	<i>0.48</i>	<i>0.47</i>	<i>0.46</i>	<i>0.46</i>	<i>0.46</i>	<i>0.45</i>

^aItalic values show the best performing correction method in terms of the highest N-S, respective to PSS, and the lowest MAE.

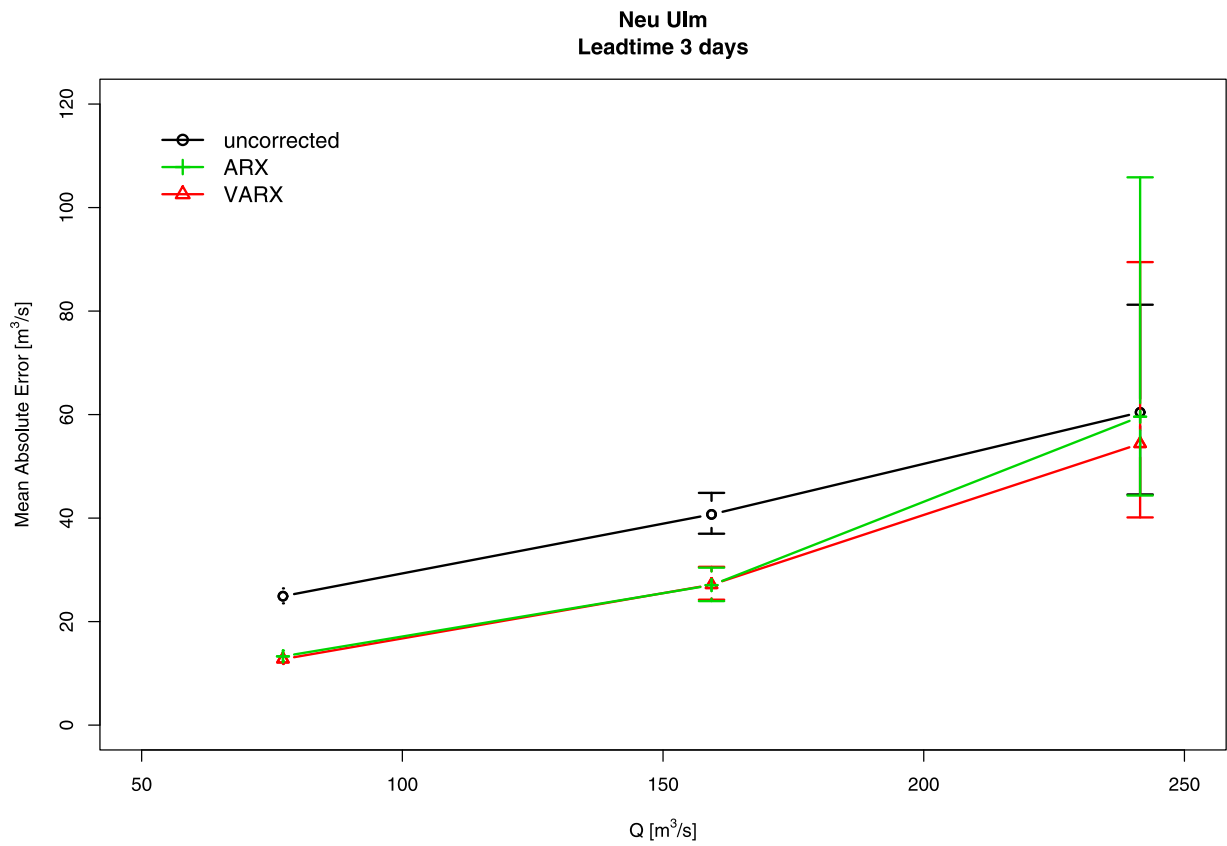


Figure 5. Mean absolute error conditional upon streamflow magnitude with adjusted bootstrap percentile intervals (95% confidence intervals) for three stations and 3 days lead time.

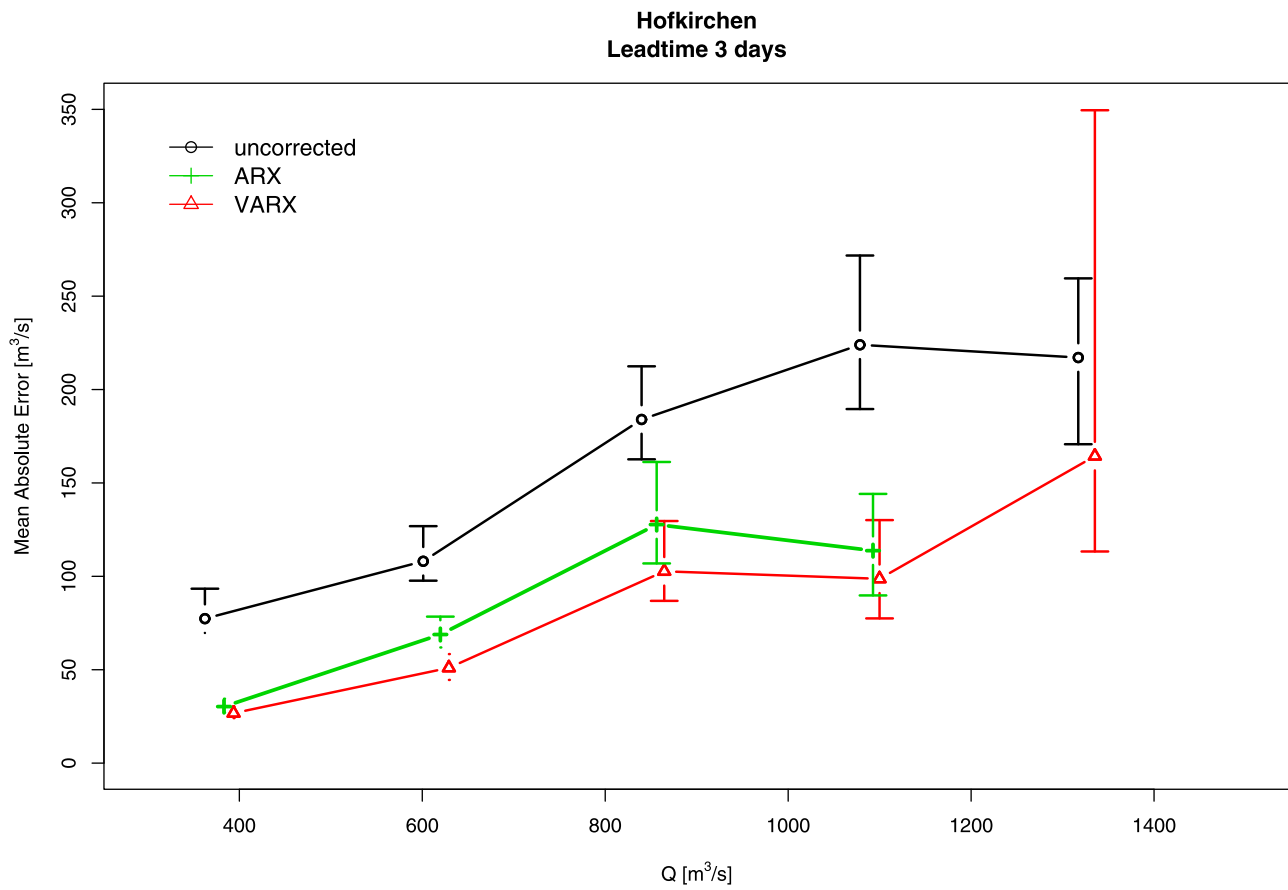


Figure 5. (continued)

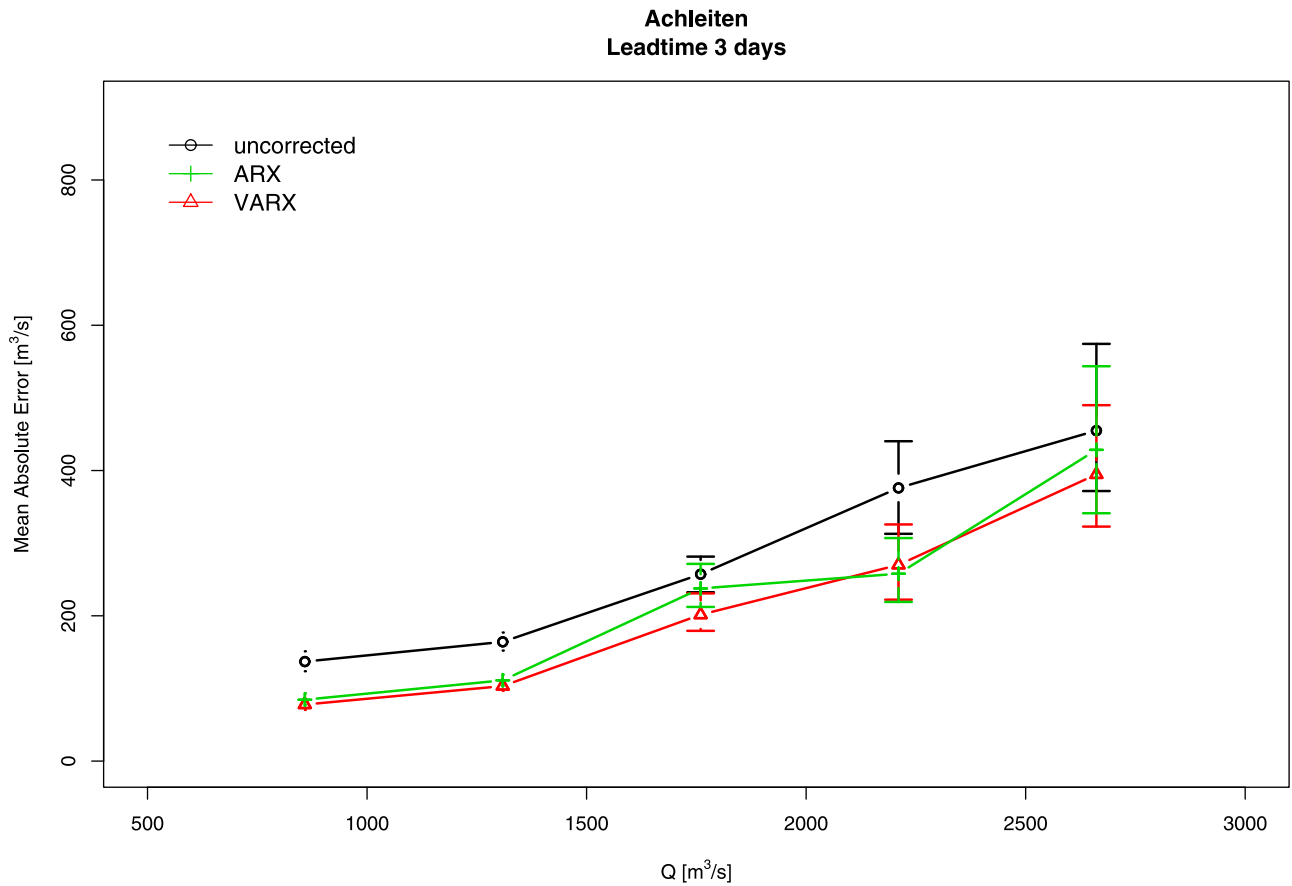


Figure 5. (continued)

Taking the observed meteorological data as input the metrics of quality comparing observation and uncorrected simulations will remain approximately constant for all lead times ranging from 1 to 10 days. The reason that the N-S coefficients for the uncorrected simulations are so low for Neu Ulm and Hofkirchen is given by the fact that the verification period includes the year 2003, which was characterized by an extreme drought resulting in extremely low flows in Central Europe, especially in the Upper Danube catchment.

[38] On the other hand the corrected simulations are updated at each time step by taking the uncorrected simulations for the next 10 days as exogenous input and by forecasting the next 10 time steps by applying the fitted autoregressive model, without applying the wavelet transformation (ARX), and with application of the wavelet transformation (VARX). Therefore the quality of the predictions decreases with increasing lead time as would be

expected, but both models (ARX and VARX) are superior to the uncorrected model and to the simple AR model for all lead times greater than one day. The “one step ahead” prediction (i.e., lead time of one day) shows almost the same result for all three updating models. It is obvious that the MAEs show the same behavior as the N-S coefficients, but it is interesting to see the absolute differences and the significant improvements in real values.

[39] Evaluation of the quality of the updating procedures based on the decomposed N-S coefficients r , α and β turned out to be rather difficult (at least for the three stations analyzed), because the differences are sometimes marginal and due to the dependence structure between the coefficients, where a possible gain of one coefficient (e.g., α) is compensated by the opposite effect of another coefficient (e.g., r or β).

Table 3. Mean Absolute Error (MAE), Nash-Sutcliffe (N-S), and Decomposed N-S (Into Measures of Correlation r , Variability α , and Bias β) for the Calibration Period^a

	Neu Ulm				Hofkirchen				Achleiten			
	Uncorrected	VARX	ARX	AR	Uncorrected	VARX	ARX	AR	Uncorrected	VARX	ARX	AR
MAE	30.24	<i>12.18</i>	12.29	13.21	132.35	<i>34.90</i>	37.32	41.90	384.68	<i>79.59</i>	84.07	86.84
N-S	0.80	0.93	0.93	0.92	0.82	0.97	0.97	0.97	0.52	0.95	0.95	0.94
r	0.92	0.97	0.97	0.96	0.92	0.99	0.99	0.98	0.92	0.98	0.97	0.97
α	1.14	0.96	0.97	1.06	0.86	0.99	0.99	1.03	0.96	0.99	0.99	1.00
β	0.07	0.02	0.01	0.01	0.13	0.01	0.01	0.00	0.58	0.01	0.01	0.01

^aItalic values show the best performing correction method in terms of the lowest MAE.

Table 4. Results of the Decomposition of the Nash-Sutcliffe Coefficient for the Validation Period at Neu Ulm Station

Measure	Method	Lead Time (days)									
		1	2	3	4	5	6	7	8	9	10
r	uncorrected	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.68	0.68	0.68
	VARX	0.97	0.94	0.91	0.89	0.88	0.87	0.85	0.84	0.83	0.82
	ARX	0.97	0.94	0.91	0.88	0.86	0.84	0.83	0.81	0.79	0.77
	AR	0.96	0.93	0.89	0.87	0.84	0.82	0.80	0.78	0.76	0.75
α	uncorrected	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.88	0.88
	VARX	0.93	0.88	0.85	0.83	0.82	0.81	0.80	0.79	0.79	0.78
	ARX	0.92	0.88	0.84	0.81	0.79	0.78	0.77	0.76	0.76	0.75
	AR	0.95	0.91	0.89	0.87	0.86	0.86	0.86	0.86	0.86	0.86
β	uncorrected	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.07
	VARX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
	ARX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
	AR	0.03	0.05	0.07	0.09	0.10	0.12	0.13	0.14	0.15	0.16

[40] Nonetheless the decomposition indicates for all three stations that the direct updating methods of VARX and ARX (see also Appendix A) are superior in reproducing the timing and the shape of the hydrographs as measured by the correlation r and in removing the bias for lead times greater than one day. Since the relative variability as measured by α keeps approximately constant for all lead times for the uncorrected simulation, it will be the same for the AR model as a result of the indirect updating procedure, at least after the first time steps. Therefore, if the model is able to reproduce the daily fluctuation realistically, α will be closer to the ideal value of one for the uncorrected and the AR model output. For example at station Neu Ulm only at the very first time steps the variability is better reproduced by the direct approach than by the uncorrected and the AR model, because thereafter the predictions of VARX and ARX will converge slowly to climatology and the variability decreases with longer lead times. Therefore the classical N-S efficiency coefficient is the preferred quality measure for comparing observed, simulated and deterministic fore-

casts in this study. For reasons of brevity only the results of the station Neu Ulm are shown in Table 4. The AR model has been excluded from further analysis because of its inferiority and only the results of the ARX and VARX models will be shown in sections 5.2 and 5.3.

5.2. Evaluation of the HUP

[41] Results of the analysis of the assumptions of *linearity*, *homoscedasticity* and *normality* are shown for the station Neu Ulm, because the size of this catchment is on the lower limit of applicability of the actual 5 km² resolution model setup of EFAS. In Figure 6 one can see that the application of the NQT removes the slight nonlinearity, whereas in Figure 7a the homoscedastic assumption of dependence is verified and Figure 7b demonstrates that the distribution function of the residuals is Gaussian. The results for the two other stations are very similar. Thus all assumptions are fulfilled.

[42] In Figures 8 and 9 the results for the evaluation of the dependence structure of the likelihood function are shown

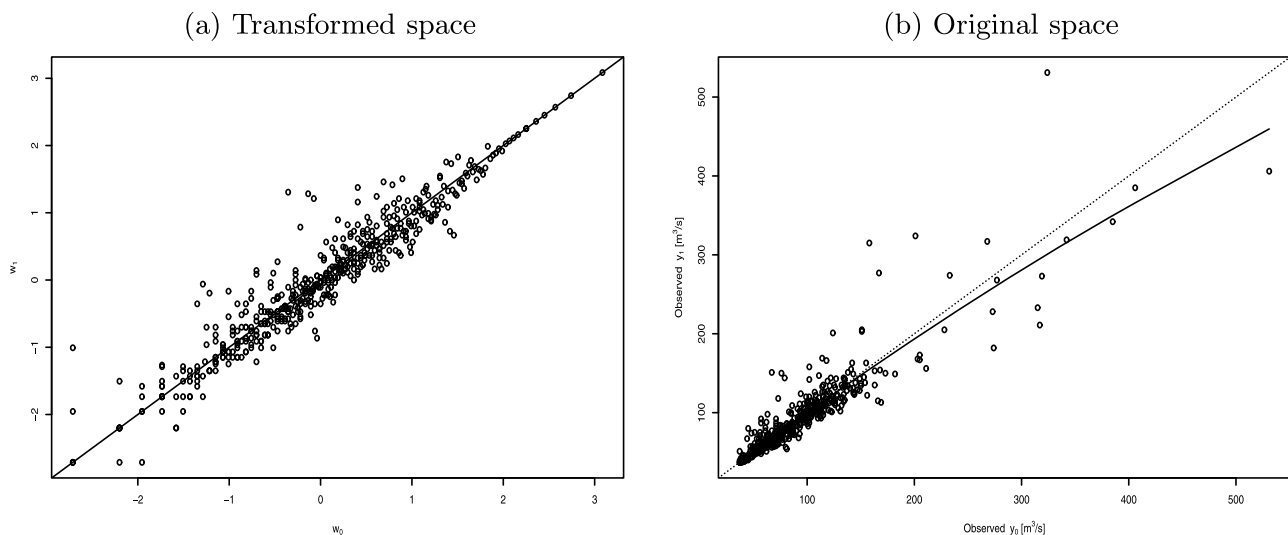


Figure 6. Dependence structure of the prior one-step transition density at station Neu Ulm indicating (a) linear dependence in the transformed space and (b) nonlinear dependence in the original space.

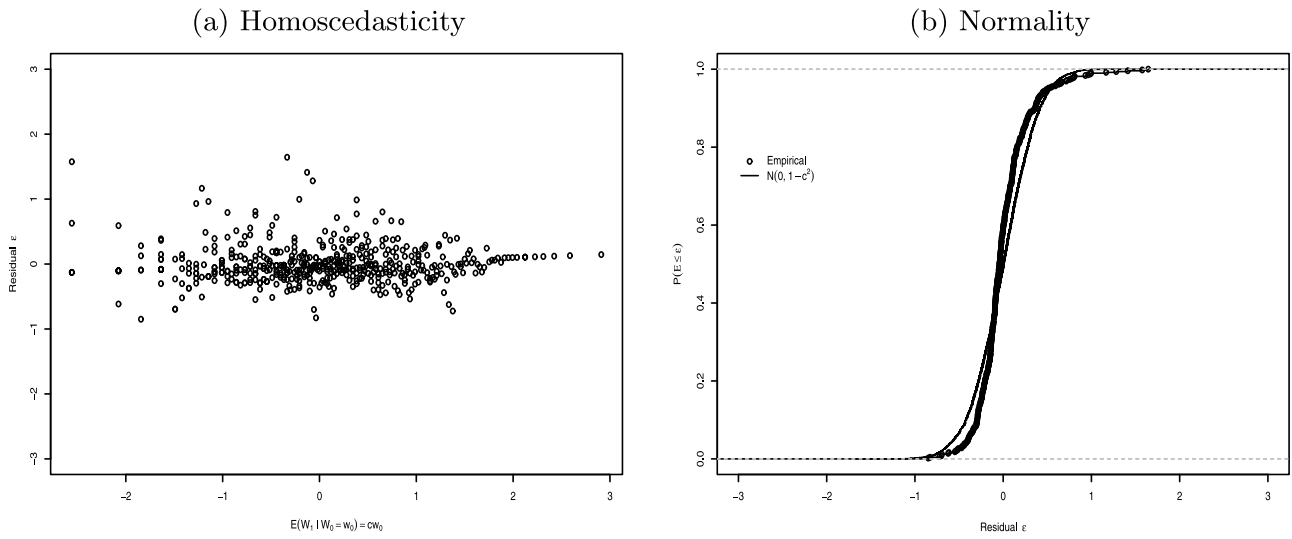


Figure 7. Validation of the normal linear model at station Neu Ulm for W_1 on W_0 : (a) homoscedasticity of dependence and (b) normality of residuals.

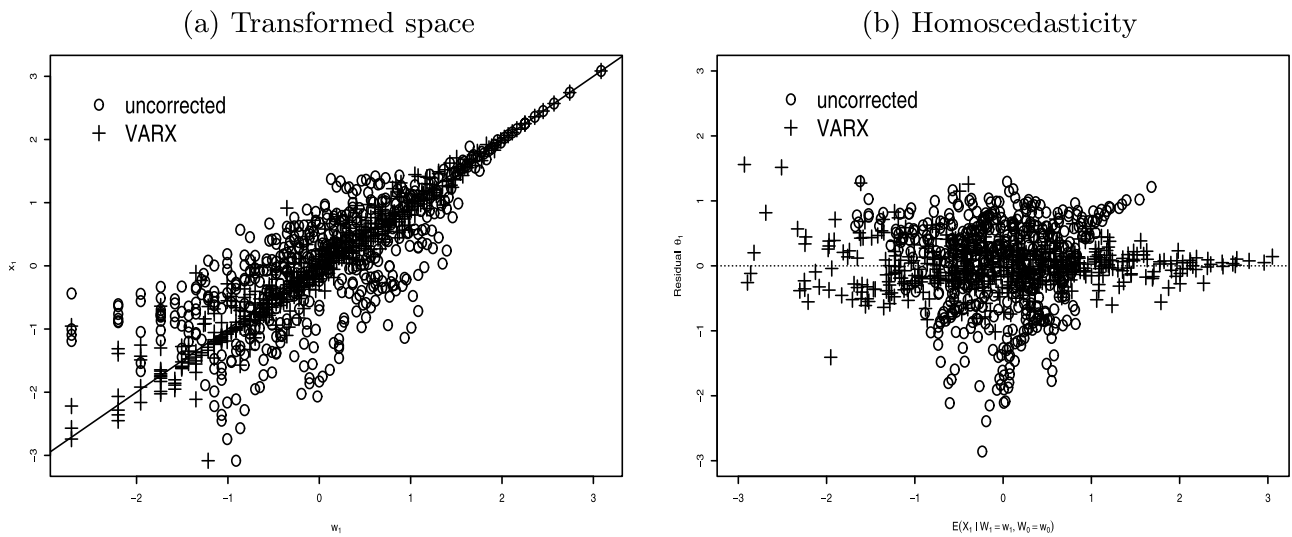


Figure 8. Analysis of the likelihood function at station Neu Ulm: (a) showing linear dependence structure between W_1 and X_1 in the transformed space and (b) validation of the homoscedasticity of the dependence.

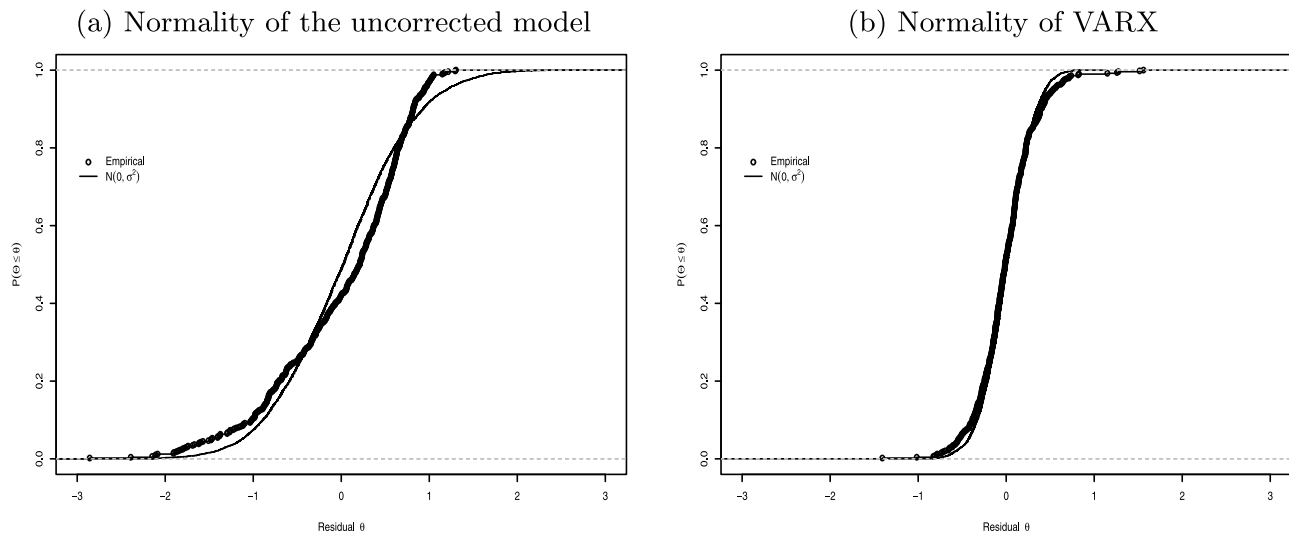


Figure 9. Validation of the dependence structure of the likelihood function at station Neu Ulm: (a) normality of the residuals of the uncorrected model and (b) normality of the residuals of the VARX.

for Neu Ulm, indicating the improvements of the VARX model in comparison to the uncorrected model, for which the application of the HUP would be problematic without modification, since some of the assumptions are not fulfilled. From this point of view the VARX model could serve as a simple but necessary data preprocessor in order to make the HUP applicable for these stations.

[43] The predictive capability of the hydrological model is evaluated in Figure 10, showing the results of the estimated likelihood parameters a and σ^2 . In addition the IS for the station Neu Ulm are displayed. The VARX model always gives the highest IS for all lead times. Only for the station Achleiten the IS of the ARX model is slightly superior to the VARX model.

5.3. Out-of-Sample Evaluation

[44] The final third of the data set is used for validation, having excluded the first part as it was used for calibrating the VARX model, and the second part as it was used for fitting the parameter of the HUP. This remaining part used for validation is referred to as the “out-of-sample” forecasting period, because all the necessary parameters for running the VARX correction and for calculating the uncertainty have been prepared and stored previously. Forecasters generally agree that forecasting methods should be assessed for accuracy using out-of-sample tests rather than goodness of fit to past data [Tashman, 2000].

[45] Thus the procedure for updating and correcting the forecast mimics the way it will be applied in realtime, i.e., running the model operationally, producing for each day a corrected forecast for the next 10 days including uncertainty ranges for each lead time. Even a modest improvement in the out-of-sample forecast error compared to that of an alternative formulation is generally taken to be strong evidence in favor of a method [Ashley, 2003].

[46] In Figure 11 the resulting predictive QQ plot for lead time 1 at station Neu Ulm is shown plotting the empirical CDF (cumulative distribution function) of z values versus

the CDF of a uniform distribution $U(0, 1)$ to evaluate forecast reliability.

[47] The results of the rank test of independence suggested that the hypothesis H_0 , namely that the values are independent and identically distributed random variables, should be rejected for the uncorrected model for the stations Neu Ulm and Hofkirchen.

[48] In Table 5 the D_{\max} values are shown demonstrating the reduced departure from uniformity by the VARX model for the station Neu Ulm and Achleiten. For the station Hofkirchen, however, the uncorrected simulation shows the lowest D_{\max} values from lead time 4 to 10 days, whereas for the first three days VARX is superior to the uncorrected and ARX methods. This clearly indicates the weakness of regarding reliability as a measure of efficiency without taking into consideration the accuracy. The higher reliability, i.e., lower D_{\max} value for the uncorrected simulation at Hofkirchen, stems from the fact that at lead times greater than three days the uncertainty of the uncorrected output is much higher in comparison to the corrected model outputs. Higher-uncertainty results from an increased spread of the probabilistic forecast, which means that the probability of an observed value attaining the predictive CDF increases resulting in uniformly distributed z values. Therefore the measure of reliability should be combined with the measure of accuracy.

[49] If accuracy (using MAE) versus reliability D_{\max} is plotted for various stations (not shown, but see values in Tables 2 and 5) then the superior behavior of the VARX model in comparison with the ARX model can be demonstrated. In such an analysis the closer the points are to zero, the better is the quality of the model. The corrected models (ARX and VARX) have a considerable clearer shift toward zero. An analysis of the uncorrected model (also not shown) using the same plotting technique demonstrates that the accuracy of the model is independent of the lead time given the observed meteorological input data. On the other hand the reliability increases (i.e., the D_{\max} decreases) with lead time as forecasts increasingly become more similar to

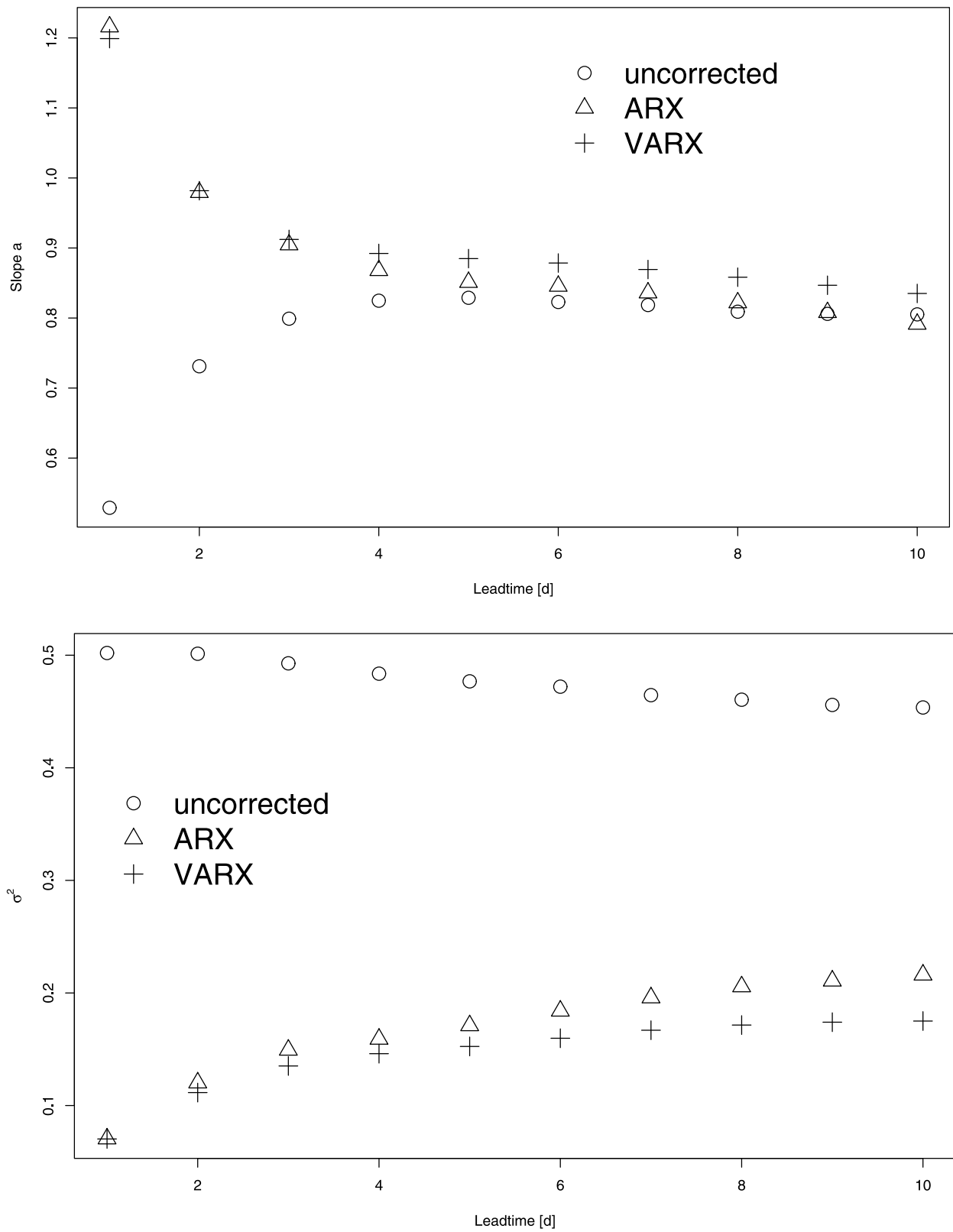


Figure 10. Comparison of the output signal, noise, and informativeness score for uncorrected, the ARX, and the VARX model at station Neu Ulm.

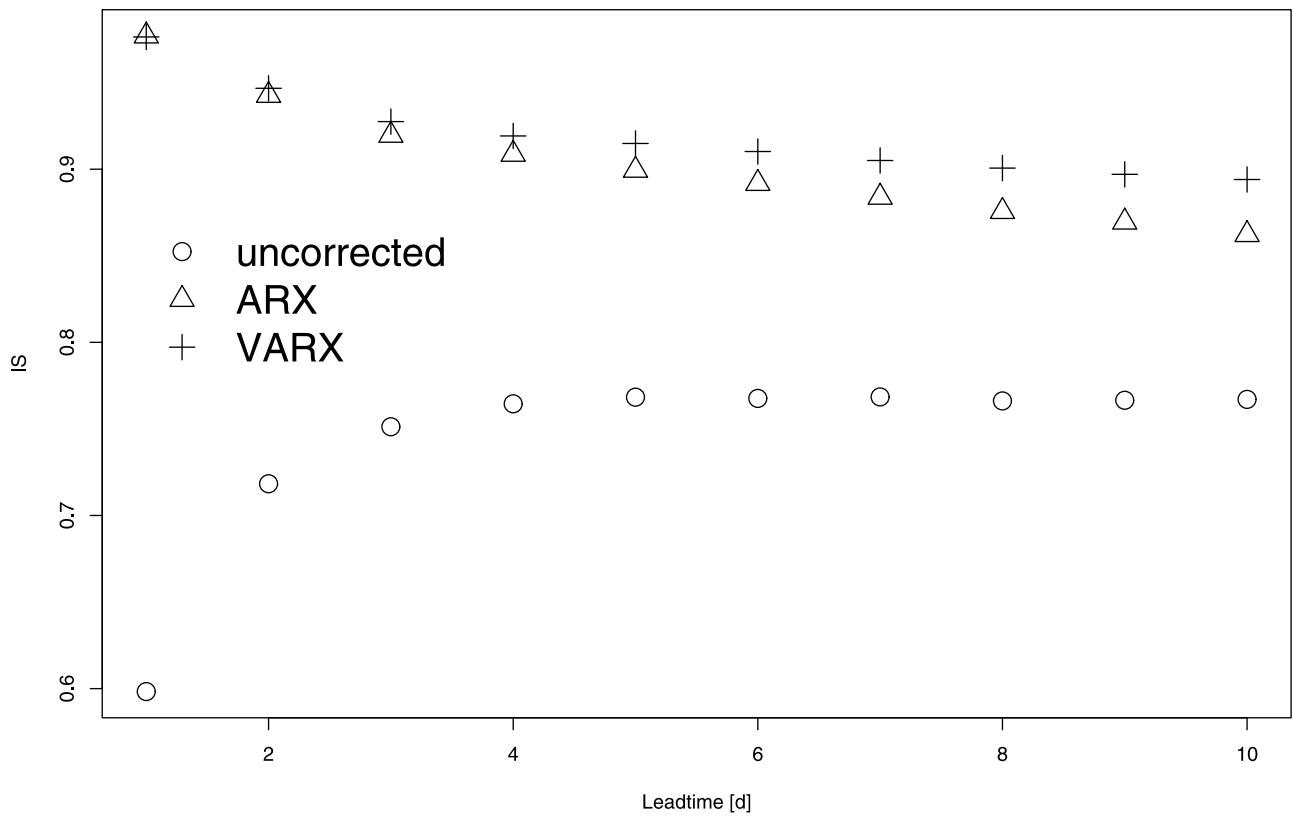


Figure 10. (continued)

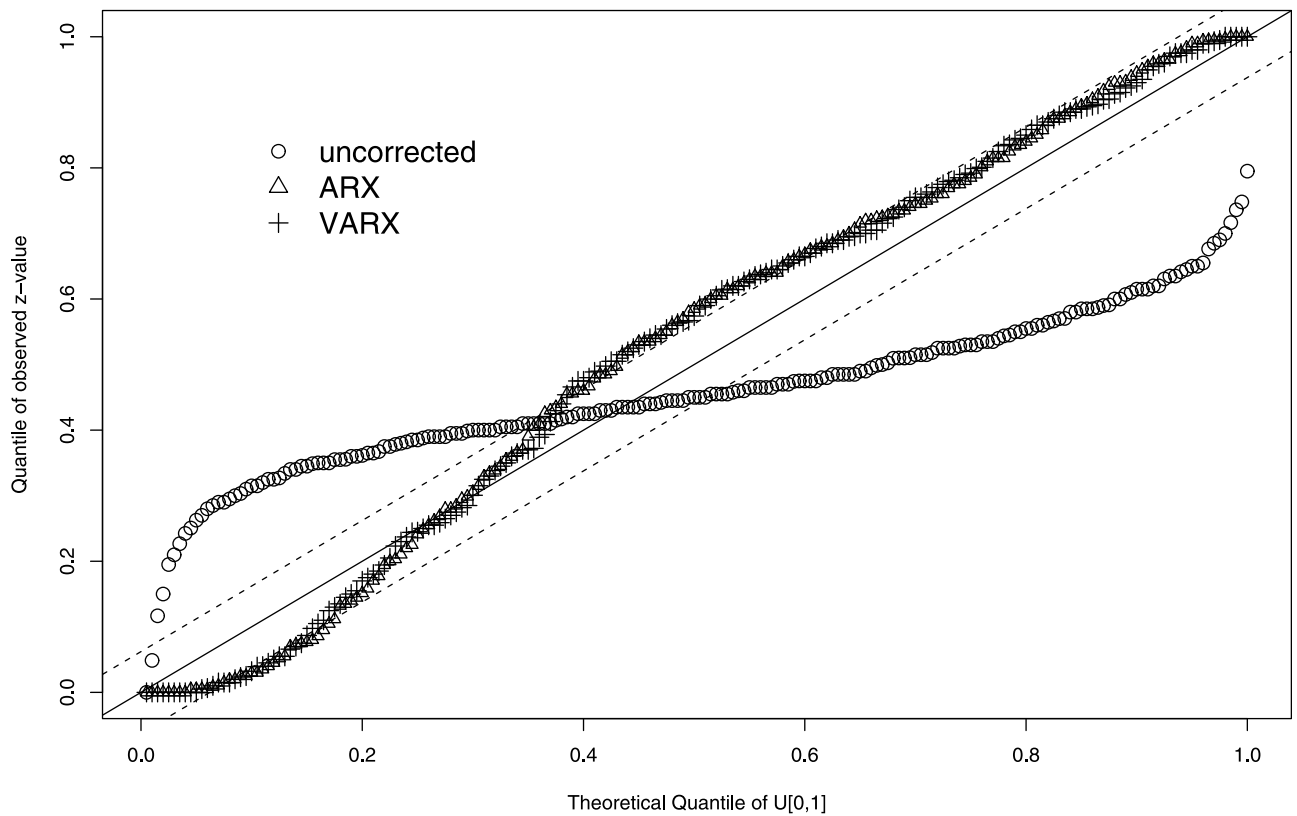


Figure 11. Predictive QQ plot for lead time 1 day at the station Neu Ulm. The dashed lines show the Kolmogorov 5% significance bands.

Table 5. Maximum Absolute Distance D_{\max}

Station	Model	Lead Time (days)									
		1	2	3	4	5	6	7	8	9	10
Neu Ulm	uncorrected	0.35	0.30	0.28	0.27	0.28	0.28	0.27	0.28	0.28	0.28
	ARX	0.23	0.21	0.21	0.22	0.23	0.23	0.23	0.23	0.26	0.26
	VARX	0.20	0.17	0.17	0.16	0.16	0.16	0.18	0.18	0.19	0.19
Hofkirchen	uncorrected	0.34	0.29	0.21	0.14	0.09	0.05	0.03	0.05	0.05	0.06
	ARX	0.32	0.21	0.20	0.22	0.23	0.25	0.26	0.27	0.26	0.26
	VARX	0.30	0.19	0.18	0.21	0.23	0.24	0.27	0.27	0.28	0.26
Achleiten	uncorrected	0.22	0.17	0.15	0.14	0.15	0.15	0.15	0.16	0.15	0.16
	ARX	0.12	0.12	0.11	0.12	0.12	0.12	0.13	0.14	0.13	0.13
	VARX	0.12	0.12	0.10	0.09	0.08	0.06	0.07	0.08	0.08	0.09

climatology. Such an analysis reveals (see also Tables 2 and 3) that for all stations the VARX model is closer to the origin, however with different shapes depending on the lead time and on the station. At the station Neu Ulm the quality of the VARX model is high for lead times up to 6 days and then decreases steadily, whereas the quality of the ARX model starts to decrease already at lead time 2. At the station Hofkirchen the superior quality of the VARX model is driven by the higher reliability for the lead times from one to three days, whereas after day four the better quality results from the higher accuracy of the prediction. At station Achleiten for the first two lead times the quality of the ARX and VARX model are similar, but for the higher lead times the VARX model is superior because of the much higher reliability. Combining the reliability and the accuracy demonstrates the clear superiority of the VARX model.

6. Conclusions

[50] In this paper the improvements, by error analysis, correction and predictive uncertainty estimation, of the efficiency of medium range streamflow predictions of the European Flood Alert System are investigated. In a first step three different methods of error correction and updating are compared using three methods: a simple autoregressive model; an autoregressive model with exogenous (ARX) input; and a method based on wavelet transforms. For the latter method the signal is decomposed into wavelets and then a vector-ARX (VARX) updating procedure is applied simultaneously. Model uncertainties are captured using the Hydrological Uncertainty Processor (HUP). The HUP has been used to derive efficiently the conditional probability distribution (CDF) of the future observed quantity given the available sample of model predictions. The comparison and validity of the method is demonstrated for three stations along the River Danube. It can be stated that (1) VARX is an essential and effective method for data preprocessing in order to meet all assumptions of the HUP (in particular the dependence structure of the likelihood function); (2) VARX in combination with the HUP leads to a better informativeness score (an improved measure of the signal-to-noise ratio) than using the ARX as preprocessor for the HUP; (3) the VARX approach is generally superior to the ARX, in particular with increasing lead time and beyond catchment concentration times, with regard to the Nash-Sutcliffe effi-

ciency coefficient and the mean absolute error within the validation period; and (4) VARX corrected simulations demonstrate a higher reliability in combination with a higher accuracy.

[51] In summary, wavelet decomposition is an excellent way to provide the detailed model error at different levels in order to estimate more precisely the (unobserved) state variables. The methodology applied in this paper is essential for deriving more reliable and accurate medium range forecasts. Future research will be needed to test the methodology on a longer data set and with the full input uncertainties.

Appendix A: VARX, ARMAX and State-Space Models

[52] The *vector autoregressive model* (VAR) has been applied successfully in capturing the evolution and interdependencies between multiple time series, generalizing the univariate AR models [Enders, 2003]. The AR forecast error updating is a special limiting procedure of the ARX model [Shamseldin and O'Connor, 1999] and serves as a benchmark model. For economic and financial time series the VAR model has proven to be especially useful for describing the dynamic behavior and for forecasting. Since it often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models, its applicability will be tested for hydrometeorological time series.

[53] Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model. This leads to the VAR model with exogenous input (VARX), which will be tested at selected gauging stations in Europe. More details about VAR can be found for example in the work by Zivot and Wang [2006, chapter 11].

[54] All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model. Based on this feature, Sims [1980] advocates the use of VAR models as a method to estimate multivariate relationships and for the modeling and forecasting of the possible dynamics relating $k \times 1$ vector valued time series $\mathbf{x}_t = (x_{t1}, \dots, x_{tk})'$, $t = 0 \pm 1, \pm 2, \dots$

[55] The extension of the univariate autoregressive model to the multivariate AR model is straight forward [Shumway

and Stoffer, 2006]. The first-order vector autoregressive model, VAR(1), is given by

$$\mathbf{x}_t = \alpha + \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, \quad (\text{A1})$$

where Φ is a $k \times k$ transition matrix expressing the dependence of \mathbf{x}_t on \mathbf{x}_{t-1} . The vector white noise process \mathbf{w}_t is assumed to be multivariate normal with mean-zero and covariance matrix $E(\mathbf{w}_t \mathbf{w}_t') = \Sigma_w$. A $k \times 1$ vector-valued time series \mathbf{x}_t for $t = 0 \pm 1, \pm 2, \dots$, is said to be VARMA(p, q) if \mathbf{x}_t is stationary and

$$\mathbf{x}_t = \alpha + \Phi_1 \mathbf{x}_{t-1} + \dots + \Phi_p \mathbf{x}_{t-p} + \mathbf{w}_t + \Theta_1 \mathbf{w}_{t-1} + \dots + \Theta_q \mathbf{w}_{t-q}, \quad (\text{A2})$$

with $\Phi_p \neq 0, \Theta_q \neq 0$, and the coefficient matrices $\Phi_j, j = 1, \dots, p$ and $\Theta_j, j = 1, \dots, q$ are $p \times p$ matrices. The constant α of the Vector Autoregressive Moving Average (VARMA) model in equation A2 can be generalized to include a fixed $r \times 1$ vector of inputs, \mathbf{u}_t , which results in the vector ARMAX model

$$\mathbf{x}_t = \Gamma \mathbf{u}_t + \sum_{j=1}^p \Phi_j \mathbf{x}_{t-j} + \sum_{k=1}^q \Theta_k \mathbf{w}_{t-k} + \mathbf{w}_t, \quad (\text{A3})$$

where Γ is a $p \times r$ parameter matrix and the exogenous vector process is denoted by \mathbf{u}_t .

[56] The VARX model (i.e., $q = 0$ in equation (A3)) is used for fitting of observed and simulated series of wavelet coefficients and the method of least squares can be applied for fitting the model for a set of possible time lags. Additionally a state-space model reduction procedure based on singular value decomposition [Mittnik, 1989; Aoki and Havenner, 1986] is applied to select the best balanced model, which is described as the brute force technique (BFT) in the work by Gilbert [1995]. Maximum likelihood estimation (MLE) would have been an alternative estimation procedure, but in the case of multivariate time series models MLE can be computationally expensive and unstable, even when a good structure and good starting parameter values are known. This is especially true for state-space models.

[57] In the general case of vector ARMA or ARMAX models, forecasts and their mean square prediction error can be obtained by using the state-space formulation of the model and the Kalman filter [Shumway and Stoffer, 2006]. In its basic form the state-space model, sometimes called dynamic linear model (DLM) also, employs a VAR(1) as the state equation

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t. \quad (\text{A4})$$

Furthermore in the state-space model approach it is assumed that the state vector \mathbf{x}_t is not observed directly, but only a linearly transformed version of it with superimposed noise,

$$\mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{v}_t, \quad (\text{A5})$$

where A_t is a $q \times p$ measurement or observation matrix and equation (A5) is called the observation equation.

[58] The VARX model in state-space form is given in equations (4) and (5). More details about DLM and State-space models in general can be found, e.g., in the work by Stoffer and Wall [2004], Petris et al. [2009], and Shumway

and Stoffer [2006]. For further definitions concerning hydrological applications of the state-space approach see Szöllösi-Nagy et al. [1997].

[59] The main difference between the AR and (V)ARX model is that the former indirectly produces the updated discharge providing estimates of the forecast error, which is added to the simulation, whereas the ARX and VARX models produce directly the updated discharge. Further details about the AR updating model and the estimation of multiple lead time forecast error time series can be found in the work by Xiong and O'Connor [2002].

[60] **Acknowledgments.** The authors wish to thank the Deutsche Wetterdienst, the European Centre for Medium-Range Weather Forecasts, the Bavarian Environment Agency, and the JRC's Institute for Protection and Security of the Citizen for data and information and especially Hannah Cloke for helpful comments. Finally, the authors gratefully acknowledge the support of all staff of the JRC's Institute for Environment and Sustainability, Land Management and Natural Hazards Unit, FLOODS Action.

References

- Aoki, M., and A. Havenner (1986), Approximate state space models of some vector-valued macroeconomic time series for cross-country comparisons, *J. Econ. Dyn. Control*, 10(1–2), 149–155.
- Ashkar, F., and C. N. Tatsambon (2007), Revisiting some estimation methods for the generalized Pareto distribution, *J. Hydrol.*, 346(3–4), 136–143.
- Ashley, R. (2003), Statistically significant forecasting improvements: How much out-of-sample data is likely necessary?, *Int. J. Forecasting*, 19(2), 229–239.
- Bartholmes, J. C., J. Thielen, M. H. Ramos, and S. Gentilini (2009), The European Flood Alert System EFAS—Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, *Hydrol. Earth Syst. Sci.*, 13(2), 141–153.
- Benaouda, D., F. Murtagh, J.-L. Starck, and O. Renaud (2006), Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting, *Neurocomputing*, 70(1–3), 139–154.
- Berkowitz, J. (2001), Testing density forecasts, with applications to risk management, *J. Bus. Econ. Stat.*, 19(4), 465–474.
- Bogner, K., and M. Kalas (2008), Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the upper Danube catchment, *Atmos. Sci. Lett.*, 9(2), 95–102.
- Bordeianu, C., R. Landau, and M. Paez (2009), Wavelet analyses and applications, *Eur. J. Phys.*, 30(5), 1049–1062.
- Briggs, W., and R. Levine (1997), Wavelets and field forecast verification, *Mon. Weather Rev.*, 125(6), 1329–1373.
- Brockwell, P. J., and R. A. Davis (2002), *Introduction to Time Series and Forecasting*, 2nd ed., Springer, New York.
- Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, 298(1–4), 242–266.
- Carmona, R., W. L. Hwang, and B. Torresani (1998), *Practical Time-Frequency Analysis: Gabor and Wavelet Transforms With an Implementation in S*, Academic, New York.
- Chou, C.-M., and R.-Y. Wang (2004), Application of wavelet-based multi-model Kalman filters to real-time flood forecasting, *Hydrol. Processes*, 18(5), 987–1008.
- Cloke, H., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, 375(3–4), 613–626.
- Conover, W. (1999), *Practical Nonparametric Statistics*, 3rd ed., Wiley, New York.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, CBMS-NSF Reg. Conf. Ser. Appl. Math., vol. 61, Soc. for Ind. and Appl. Math., Philadelphia, Pa.
- Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Applications*, Cambridge Univ. Press, Cambridge, U. K.
- Dawid, A. (1984), Statistical theory: The prequential approach, *J. R. Stat. Soc., Ser. A*, 147(2), 278–292.
- De Groot, M. (1970), *Optimal Statistical Decisions*, McGraw Hill, New York.
- DiCiccio, T., and B. Efron (1996), Bootstrap confidence intervals, *Stat. Sci.*, 11(3), 189–228.

- Diebold, F., T. Gunther, and A. Tay (1998), Evaluating density forecasts with applications to financial risk management, *Int. Econ. Rev.*, 39(4), 863–883.
- Duan, Q., and S. Sorooshian (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015–1031.
- Dutilleul, P. (1989), An implementation of the “algorithme a trous” to compute the wavelet transform, in *Wavelets: Time-Frequency Methods and Phase Space*, edited by J. M. Combes, A. Grossmann, and P. Tchamitchian, pp. 298–304, Springer, New York.
- Enders, W. (2003), *Applied Econometric Time Series*, 2nd ed., 433 pp., John Wiley, Hoboken, N. J.
- Feyen, L., J. A. Vrugt, B. O. Nuallain, J. van der Knijff, and A. D. Roo (2007), Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model, *J. Hydrol.*, 332(3–4), 276–289.
- Feyen, L., M. Kalas, and J. Vrugt (2008), Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization, *Hydrol. Sci. J.*, 53(2), 293–308.
- Freedman, D. (1981), Bootstrapping regression models, *Ann. Stat.*, 9, 1218–1228.
- Georgakakos, K. P., and G. F. Smith (1990), On improved hydrologic forecasting—Results from a WMO real-time forecasting experiment, *J. Hydrol.*, 114(1–2), 17–45.
- Gilbert, P. (1995), Combining VAR estimation and state space model reduction for simple good predictions, *J. Forecasting*, 14, 229–250.
- Glahn, H., and D. Lowry (1972), The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteorol.*, 11, 1203–1211.
- Gneiting, T., and A. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, 102(477), 359–378.
- Gneiting, T., A. Raftery, A. Westveld III, and T. Goldman (2005), Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, *Mon. Weather Rev.*, 133(5), 1098–1118.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377(1–2), 80–91.
- He, Y., F. Wetterhall, H. L. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor (2009), Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, *Meteorol. Appl.*, 16(1), 91–101.
- Herr, H. D., and R. Krzysztofowicz (2010), Bayesian ensemble forecast of river stages and ensemble size requirements, *J. Hydrol.*, 387(3–4), 151–164.
- Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecasting*, 15(5), 559–570.
- Kachroo, R., and G. Liang (1992), River flow forecasting: Part 2. Algebraic development of linear modelling techniques, *J. Hydrol.*, 133(1–2), 17–40.
- Kelly, K., and R. Krzysztofowicz (1994), Probability distributions for flood warning systems, *Water Resour. Res.*, 30(4), 1145–1152.
- Kelly, K., and R. Krzysztofowicz (1997), A bivariate meta-Gaussian density for use in hydrology, *Stochastic Hydrol. Hydraul.*, 11(1), 17–31.
- Kelly, K., and R. Krzysztofowicz (2000), Precipitation uncertainty processor for probabilistic river stage forecasting, *Water Resour. Res.*, 36(9), 2643–2653.
- Krzysztofowicz, R. (1983), Why should a forecaster and a decision maker use Bayes theorem, *Water Resour. Res.*, 19(2), 327–336.
- Krzysztofowicz, R. (1992), Bayesian correlation score: A utilitarian measure of forecast skill, *Mon. Weather Rev.*, 120(1), 208–219.
- Krzysztofowicz, R. (1997), Transformation and normalization of variates with specified distributions, *J. Hydrol.*, 197(1–4), 286–292.
- Krzysztofowicz, R. (1999a), Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, 35(9), 2739–2750.
- Krzysztofowicz, R. (1999b), Bayesian forecasting via deterministic model, *Risk Anal.*, 19(4), 739–749.
- Krzysztofowicz, R. (2001a), The case for probabilistic forecasting in hydrology, *J. Hydrol.*, 249(1–4), 2–9.
- Krzysztofowicz, R. (2001b), Integrator of uncertainties for probabilistic river stage forecasting: Precipitation-dependent model, *J. Hydrol.*, 249(1–4), 69–85.
- Krzysztofowicz, R., and W. Evans (2008), Probabilistic forecasts from the national digital forecasts database, *Weather Forecasting*, 23(2), 270–289.
- Krzysztofowicz, R., and K. Kelly (2000), Hydrologic uncertainty processor for probabilistic river stage forecasting, *Water Resour. Res.*, 36(11), 3265–3277.
- Krzysztofowicz, R., and C. J. Maranzano (2004), Hydrologic uncertainty processor for probabilistic stage transition forecasting, *J. Hydrol.*, 293(1–4), 57–73.
- Krzysztofowicz, R., and C. J. Maranzano (2006), Bayesian processor of output: Probabilistic quantitative precipitation forecast, working paper, Univ. of Va., Charlottesville.
- Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, 11(4), 1267–1277.
- Lane, S. N. (2007), Assessment of rainfall-runoff models based upon wavelet analysis, *Hydrol. Processes*, 21(5), 586–607.
- Liu, Y., J. Brown, J. Demargne, and D.-J. Seo (2011), A wavelet-based approach to assessing timing errors in hydrologic predictions, *J. Hydrol.*, 397(3–4), 210–224.
- Mallat, S. G. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, 11, 674–693.
- Mittnik, S. (1989), Multivariate time series analysis with state space models, *Comput. Math. Appl.*, 17(8–9), 1189–1201.
- Mittnik, S. (1990), Forecasting with balanced state space representations of multivariate distributed lag models, *J. Forecasting*, 9, 207–218.
- Mo, X., F. Pappenberger, K. Beven, S. Liu, A. De Roo, and Z. Lin (2006), Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model, *Hydrol. Sci. J.*, 51(1), 45–65.
- Murphy, A. H. (1988), Skill scores based on the mean square error and their relationships to the correlation coefficient, *Mon. Weather Rev.*, 116(12), 2417–2424.
- Murphy, A. H., and H. Daan (1985), Forecast evaluation, in *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, edited by A. H. Murphy and R. W. Katz, pp. 379–437, Westview, Boulder, Colo.
- Murphy, A. H., and R. Winkler (1987), A general framework for forecast verification, *Mon. Weather Rev.*, 115(7), 1330–1338.
- Murphy, A. H., and R. L. Winkler (1992), Diagnostic verification of probability forecasts, *Int. J. Forecasting*, 7(4), 435–455.
- Murphy, A. H., B. Brow, and Y.-S. Chen (1989), Diagnostic verification of temperature forecasts, *Weather Forecasting*, 4, 485–501.
- Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, 10(3), 282–290.
- Nason, G. P. (2008), *Wavelet Methods in Statistics With R*, 1st ed., Springer, New York.
- Noceti, P., J. Smith, and S. Hodges (2003), An evaluation of tests of distributional forecasts, *J. Forecasting*, 22(6–7), 447–455.
- O’Connell, P., and R. Clarke (1981), Adaptive hydrological forecasting—A review, *Hydrol. Sci. Bull.*, 26(2), 179–205.
- Percival, D., and A. Walden (2000), *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press, Cambridge, U. K.
- Petris, G., S. Petrone, and P. Campagnoli (2009), *Dynamic Linear Models With R*, Springer, Dordrecht, Netherlands.
- Pickands, J., III (1975), Statistical inference using extreme order statistics, *Ann. Stat.*, 3, 119–131.
- Qian, T., M. Vai, and Y. Xu (Eds.) (2007), *Wavelet Analysis and Applications*, Springer, Basel, Switzerland.
- Refsgaard, J. C. (1997), Validation and intercomparison of different updating procedures for real-time forecasting, *Nordic Hydrol.*, 28, 65–84.
- Regonda, S. K., B. Sivakumar, and A. Jain (2004), Temporal scaling in river flow: Can it be chaotic?, *Hydrol. Sci. J.*, 49(3), 373–385.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Renaud, O., J.-L. Starck, and F. Murtagh (2003), Prediction based on a multiscale decomposition, *Int. J. Wavelets Multiresolut. Inf. Process.*, 1(2), 217–232.
- Rosenblatt, M. (1952), Remarks on a multivariate transformation, *Ann. Math. Stat.*, 23, 470–472.
- Rougier, J. (2007), Probabilistic inference for future climate using an ensemble of climate model evaluations, *Climat. Change*, 81, 247–264.
- Saito, N., and G. Beylkin (1993), Multiresolution representations using the auto-correlation functions of compactly supported wavelets, *IEEE Trans. Signal Process.*, 41(12), 3584–3590.
- Shamseldin, A., and K. O’Connor (1999), A real-time combination method for the outputs of different rainfall-runoff models, *Hydrol. Sci. J.*, 44(6), 895–912.
- Shumway, R. H., and D. S. Stoffer (2006), *Time Series Analysis and Its Applications: With R Examples*, 2nd ed., Springer, New York.
- Sims, C. (1980), Macroeconomics and reality, *Econometrica*, 48(1), 1–48.

- Stankovic, R., and B. Falkowski (2003), The Haar wavelet transform: Its status and achievements, *Comput. Electr. Eng.*, 29(1), 25–44.
- Stoffer, D., and K. Wall (2004), Resampling in state space models, in *State Space and Unobserved Component Models*, edited by A. Harvey, S. Koopman, and N. Shephard, pp. 171–202, Cambridge Univ. Press, Cambridge, U. K.
- Szöllösi-Nagy, A., E. Todini, and E. F. Wood (1997), State-space model for realtime forecasting of hydrological time-series, *J. Hydrol. Sci.*, 4, 1–11.
- Tashman, L. J. (2000), Out-of-sample tests of forecasting accuracy: An analysis and review, *Int. J. Forecasting*, 16(4), 437–450.
- Thielen, J., J. Bartholmes, M.-H. Ramos, and A. de Roo (2009), The European Flood Alert System—Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, 13(2), 125–140.
- Thiemig, V., F. Pappenberger, J. Thielen, H. Gadain, A. de Roo, K. Bodis, M. Del Medico, and F. Muthusi (2010), Ensemble flood forecasting in Africa: A feasibility study in the Juba-Shabelle river basin, *Atmos. Sci. Lett.*, 11(2), 123–131.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Todini, E. (2008), A model conditional processor to assess predictive uncertainty in flood forecasting, *Int. J. River Basin Manage.*, 6(2), 123–137.
- Todini, E. (2009), Predictive uncertainty assessment in real time flood forecasting, in *Uncertainties in Environmental Modelling and Consequences for Policy Making*, edited by P. C. Baveye, M. Laba, and J. Mysiak, pp. 205–228, Springer, Dordrecht, Netherlands.
- Van Der Knijff, J. M., J. Younis, and A. P. J. De Roo (2010), LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.*, 24(2), 189–212.
- Wilks, D. (2009), Extending logistic regression to provide full-probability-distribution MOS forecasts, *Meteorol. Appl.*, 16(3), 361–368.
- World Meteorological Organization (1992), *Simulated Real-Time Intercomparison of Hydrological Models*, Oper. Hydrol. Rep. 38, Geneva, Switzerland.
- Xiong, L., and K. O'Connor (2002), Comparison of four updating models for real-time river flow forecasting, *Hydrol. Sci. J.*, 47(4), 621–640.
- Zheng, G., J.-L. Starck, J. Campbell, and F. Murtagh (1999), The wavelet transform for filtering financial data streams, *J. Comput. Intell. Finance*, 7(3), 18–35.
- Zivot, E., and J. Wang (2006), *Modeling Financial Time Series With S-PLUS*, 2nd ed., Springer, Berlin.

K. Bogner, Land Management and Natural Hazards Unit, Institute for Environment and Sustainability, European Commission Joint Research Centre, Via E. Fermi 2749, Ispra, I-21027 Varese, Italy. (konrad.bogner@jrc.ec.europa.eu)

F. Pappenberger, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK.