Estimating the uncertainty of hydrological predictions through data-driven resampling techniques

Anna E. Sikorska[1], Alberto Montanari [2] and Demetris Koutsoyiannis[3]

Short title:

RESAMPLING TECHNIQUES FOR UNCERTAINTY ESTIMATION

[1] Visiting researcher. Department DICAM, University of Bologna, Via del Risorgimento 2, I-40136 Bologna, Italy; Research hydrologist. Department of Hydraulic Engineering, Warsaw University of Life Sciences-SGGW, Nowoursynowska Street 166, PL-02-787 Warsaw, Poland (corresponding author). E-mail: anna_sikorska@sggw.pl

[2] Professor. Department DICAM, University of Bologna, Via del Risorgimento 2, I-40136 Bologna, Italy. E-mail: alberto.montanari@unibo.it

[3] Professor. Department of Water Resources and Environmental Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR 157 80, Zographou, Greece. E-mail: dk@itia.ntua.gr

Abstract

Estimating the uncertainty of hydrological models remains a relevant challenge in applied hydrology, mostly because it is not easy to parameterize the complex structure of hydrological model errors. A non-parametric technique is proposed as an alternative to parametric error models to estimate the uncertainty of hydrological predictions. Within this approach, the above uncertainty is assumed to depend on input data uncertainty, parameter uncertainty and model error, where the latter aggregates all sources of uncertainty that are not considered explicitly. Errors of hydrological models are simulated by resampling from their past realizations using a nearest neighbor approach, therefore avoiding a formal description of their statistical properties. The approach is tested using synthetic data which refer to the case study located in Italy. The results are compared with those obtained using a formal statistical technique (meta-Gaussian approach) from the same case study. Our findings prove that the nearest neighbor approach provides simplicity in application and a significant improvement in regard to the meta-Gaussian approach. Resampling techniques appear therefore to be an interesting option for uncertainty assessment in hydrology, provided that historical data are available to provide a consistent description of the model error.

Keywords: hydrological forecasting; uncertainty assessment; rainfall-runoff; flood forecasting.

Introduction

Uncertainty assessment for hydrological predictions still remains a relevant challenge in applied hydrology (Bogner and Pappenberger 2011; Mendoza et al. 2012; Montanari 2011; Sikorska et al. 2012; Tada and Beven 2012). In fact, data scarcity and limited understanding of the processes governing the water cycle, together with the difficulties and costs implied by efficient and extensive monitoring campaign, very often prevent a satisfactory assessment of the reliability of hydrological predictions. Yet, uncertainty assessment is very much relevant for estimating design variables in enginnering practices (Moretti and Montanari 2008), mitigating hydrological risks and improving water resources management policies (Koutsoyiannis 2013; Montanari et al. 2013).

The problem of uncertainty assessment is in principle reducible to estimating and integrating the main sources of uncertainty through the modeling chain. The literature proposed several contributions on the estimation of input uncertainty (Clark and Slater 2006; Cunha et al. 2012; He et al. 2011; Legleiter et al. 2011; McMillan et al. 2011; Montanari and Di Baldassarre 2013; Renard et al. 2011; Sikorska et al. 2012; Sun and Bertrand-Krajewski 2013), parameter uncertainty (Ebtehaj et al. 2010; Srikanthan et al. 2009; Vrugt and Robinson 2007), measurement errors/output uncertainty (Di Baldassarre and Montanari 2009; McMillan et al. 2010; Sikorska et al. 2013) and uncertainty in the model structure (Krzysztofowicz 2002; Montanari and Brath 2004; Montanari and Grossi 2008), but the related problems are far from being solved.

Actually, the estimation of input and parameter probability distributions for hydrological models through the above mentioned methods is often affected by the presence of model structural errors which in turn are themselves related to data errors and parameter errors. Indeed, individual contributions to the global uncertainty cannot be quantified independently, unless (1) one introduces assumptions about the nature of the individual error components or (2) observations are available that allow estimating each source of error independently (Renard et al. 2011). For instance, the latter case was examined by Winsemius et al. (2006; 2008; 2009) who used gravity and evaporation measurements to constrain parameter estimation for a rainfall-runoff model. In the former case, the assumptions that are introduced to estimate each uncertainty source separately can affect the reliability of estimations (Beven 2006; 2010; 2012). A comprehensive review of uncertainty assessment techniques and their underlying assumptions has been recently presented by Montanari (2011), while the issue of introducing assumptions in uncertainty assessment in hydrology has been further discussed by Montanari and Koutsoyiannis (2012).

In this paper we mainly focus on the description of the model error, which represents a relevant challenge for both improving the understanding of hydrological processes and consistently estimating model uncertainty (Gupta et al. 2012). Some methods implicitly account for model error, like it is done by several applications of the well known and commonly applied GLUE technique (Beven and Binley 1992; Liu et al. 2009; see also the discussion in Beven et al. 2012; Clark et al. 2011). Other methods are based on comparative experiments by using different models (Clark et al. 2008). Another possible solution is the use of multi-modeling techniques like Bayesian Model Averaging (Ajami et al. 2007; Neuman 2003). However, these latter methods require establishing a likelihood to estimate the probability of each model being correct and provide a consistent estimation only if a large sample of possible models is explored.

As for the integration of the different uncertainty sources, several methods adopt a numerical procedure. According to that, the probability distribution of the model outcome is estimated by producing several outputs, which are obtained by randomly sampling the feasible spaces of input data, parameters and model structure errors usually merged together with measurement errors (Ajami et al. 2007; Krueger et al. 2010; Liu et al. 2009; Neuman 2003).

Bayesian methods constitute a promising approach to integrate all defined sources of uncertainty and propagate them through the model in order to derive uncertainty on the predicted outputs (Sikorska et al. 2012; Yang et al. 2007). Although Bayesian methods have been demonstrated to be statistically consistent and more satisfying than other uncertainty analysis approaches (Mantovan and Todini 2006), they essentially require an explicit statistical representation for the model error through the formulation of a likelihood function. The latter can, however, be a challenging task for complex hydrological models, particularly when weak or no information is available to infer the prior information on the error model (Bulygina and Gupta 2010; Sikorska 2012). Recently, improved likelihood functions have been proposed for hydrological models (Schoups and Vrugt 2010; Pianosi and Raso 2012). However, they require the introduction of additional parameters (see below). Moreover, their application may be highly time-intensive for long time series analysis and therefore less practical in real time applications when there is little time to perform the uncertainty analysis (Shrestha et al. 2009).

Alternatively, several methods have been proposed to estimate the simulation uncertainty by directly inferring the statistical properties of the simulated data by means of data assimilation techniques (Bulygina and Gupta 2009, 2010 and 2011). This allows avoiding individual quantification of the uncertainty sources and their integration. Some of these techniques rely on inferring the statistical properties of the model error, which is assumed to represent the aggregated contribution over all uncertainty sources (Montanari and Brath 2004; Montanari and Grossi 2008).

A generalized approach (a blueprint) to carry out the uncertainty assessment for model predictions has been recently proposed by Montanari and Koutsoyiannis (2012). By admitting that uncertainty in hydrological models mainly originates from model structural inadequacy (which descends from limited knowledge and therefore is an epistemic form of uncertainty), uncertainty in the observed data (input and output observations) and inherent randomness (e.g. due to sensitive dependence on initial conditions), the basic assumption in the above blueprint lies in the recognition that randomness is an intrinsic property of hydrological processes. In this respect, randomness could be thought of as indeterminacy or inherent inability to describe the future evolution of hydrological processes deterministically. The notion of indeterminacy is used here to underline our belief that a perfect reproduction of hydrological processes at scales of practical interest will never be possible and therefore the model error is also the result of an intrinsic property rather than just model inadequacy.

According to Montanari and Koutsoyiannis (2012), input and parameter uncertainties can be estimated individually while the model error is used to represent an aggregated form of all other sources of error (epistemic or induced by inherent randomness), and in particular the model structural uncertainty. The numerical integration of the different uncertainty sources is then operated by performing several model simulations, where input data and parameters are picked up from the respective feasible spaces. The model structural uncertainty is accounted for, along with indeterminacy, by adding a random outcome from the model error. The proposed scheme is particularly appealing in that it allows implicitly accounting for inherent randomness.

However, the modeling solution proposed by Montanari and Koutsoyiannis (2012) still requires a statistical characterization of the model error for randomly sampling from the related probability distribution. In their applications, the authors used the meta-Gaussian approach (Montanari and Brath 2004) to provide a time varying representation of the probability distribution of the model error. This solution seemed to have low efficiency in interpreting the heteroscedasticity of the error itself. Thus, the statistical characterization of the model error still remains a problem (Montanari and Koutsoyiannis 2012).

In this paper, we propose an alternative solution to account for the uncertainties aggregated in the model error within the blueprint introduced by Montanari and Koutsoyiannis (2012). Namely, we propose to use a non-parametric technique to obtain error realizations from the feasible space of the past model errors without the need for their explicit statistical characterization. Thus, we employ a resampling procedure in order to retrieve sufficient information of the hydrological model behavior (and its deviation from the expected - observed values) taking advantage of sufficient historical data. In particular, we perform this resampling by randomly picking up, using a nearest neighbor (NN) approach, outcomes from past model errors. These are taken from the hydrological model simulation of historical data in validation mode. The driving variable for applying the NN technique is the hydrological model simulation at each time step itself, therefore, preserving the heteroscedasticity of the model error for different river flow regimes.

Application of the NN method in hydrology is not new. In fact, the NN method has found already implementations in a wide range of real-world settings as pattern recognition, machine learning and database querying (Liu et al. 2004; Shrestha et al. 2009; Shrestha and Solomatine 2006) and for searching a model space (Beven and Binley 1992). Karlsson and Yakowitz (1987a) have demonstrated the usefulness of the NN method to large-sample time series problems. Due to its intuitiveness, simplicity (non-parametric property) and the sound theoretical basis, it has been made also attractive to forecasters in the hydrologic field for time series predictions (Brath et al. 2002; Karlsson and Yakowitz 1987b; Koutsoyiannis et al. 2008; Toth et al. 1999).

The approach proposed in this paper is tested by using 50 years of synthetic data referred to a river basin in Italy (Secchia River). The obtained results show that NN is a very efficient solution for solving the problem of characterizing the model error in hydrological predictions for long time series when sufficient historical data are available.

Theoretical setting for uncertainty assessment

The theoretical blueprint introduced by Montanari and Koutsoyiannis (2012) relies on converting a deterministic hydrological model into a stochastic one and thus incorporating randomness into hydrological modeling. This allows estimating the probability distribution of outputs from process-based (deterministic) hydrological models. For a detailed description, the reader is referred to Montanari and Koutsoyiannis (2012). Here we provide only its brief summary relevant for this study.

The theoretical blueprint scheme can be applied to any type of model. In this paper we particularly refer to a rainfall-runoff model which can be written as

$$\boldsymbol{Q} = S(\boldsymbol{\Theta}, \boldsymbol{X}) \tag{1}$$

where $\boldsymbol{Q}$ is river flow, S is a deterministic function representing the transformation model, $\boldsymbol{X}$ is the input data vector (which may include boundary conditions) and $\boldsymbol{\Theta}$ is the model parameter vector. The formalism in eq. (1) has been given in terms of converting a single deterministic model into a stochastic one. However, a multimodel framework (for an example of application see Krueger et al. (2010)) can also be considered within the blueprint framework as discussed in detail in Montanari and Koutsoyiannis (2012).

To take uncertainty into account, one needs to convert the deterministic Eq. (1) to a stochastic relationship, which applies to probability distributions of input and output data and parameters, therefore taking the form

$$f_Q(\boldsymbol{Q}) = K_S\, f_{\Theta,X}(\boldsymbol{\Theta}, \boldsymbol{X}) \tag{2}$$

where f indicates a probability density function and K_S is a stochastic operator which depends on (but is different from) the deterministic model S. In detail, Montanari and Koutsoyiannis (2012) proved that the deterministic Eq. (2) can be converted to the stochastic form

$$f_Q(\boldsymbol{Q}) = \int_\Theta \int_X f_e\left(\boldsymbol{Q} - S(\boldsymbol{\Theta}, \boldsymbol{X}) \middle| \boldsymbol{\Theta}, \boldsymbol{X}\right) f_\Theta(\boldsymbol{\Theta}) f_X(\boldsymbol{X}) d\boldsymbol{\Theta} d\boldsymbol{X} \tag{3}$$

where $\boldsymbol{Q} - S(\boldsymbol{\Theta}, \boldsymbol{X}) = \boldsymbol{e}$ is the model error which incorporates all uncertainties that are not explicitly considered in Eq. (3), namely, the input and parameter uncertainty. These latter are quantified by the probability density functions $f_X(\boldsymbol{X})$ and $f_\Theta(\boldsymbol{\Theta})$, respectively. Therefore, the model error includes all information on model structural adequacy, which depends on model structure, scales of application and specific behaviors of the case study (Blöschl et al. 1995; Skøien et al. 2003). It is relevant to note that input, parameter and data uncertainty are assumed independent of each other (see Montanari and Koutsoyiannis (2012) for an extended discussion).

The double integral in Eq. (3) can be solved numerically by performing a simulation procedure that is structured according to the following steps: (1) random outcomes for the input data vector and the parameter vector are picked up from the related probability distributions $f_X(\boldsymbol{X})$ and $f_\Theta(\boldsymbol{\Theta})$ respectively; (2) the hydrological model is run to obtain a single simulation of the output q and (3) a random outcome e from the probability distribution of the model error $f_e(\mathbf{e})$ is added, which is sampled according to the procedure described below in the next section. By repeating the above simulation procedure a sufficient (j) number of times, we obtain a number of model outcomes $\boldsymbol{Q} = \boldsymbol{q} + \boldsymbol{e}$, from which the related probability distribution $f_Q(\boldsymbol{Q})$ can be inferred. Figure 1 shows a sketch of the simulation chain.

We may note that the proposed simulation procedure is similar to GLUE (Beven and Binley 1992; Liu et al. 2009) with the exception that GLUE rejects non-behavioral simulations by usually adopting a likelihood measure. Moreover, one may note that in many applications GLUE was used without including the random contribution of the model error in the formation of the output uncertainty.

It is important to note that the methodology proposed here relies much on data. Although probability distributions of input data, model parameter and model error could be estimated according to expert knowledge, data analysis is a fundamental requirement to properly estimate the probability distribution of the model error. Therefore, particular attention should be paid to data collection and checking (Beven and Westerberg, 2011).

Sampling from the probability distribution of the model error

The Nearest Neighbor approach

The outcomes from the model error to be plugged into the simulation procedure on the step 3, as described in the previous section, are obtained here by resampling a past realization of the model error using a nearest neighbor (NN) approach also known as the *k*-nearest neighbor algorithm. This method takes advantage of the fact that a hydrological system's behavior is encapsulated into observations and therefore the stochastic dynamics of the system can be recovered if enough data are available, under assumptions of stationarity and ergodicity. To this end, the NN algorithm (Karlsson and Yakowitz 1987a; b) is applied to represent the behavior of the system through establishing a dependency between the known real inputs into the system and the corresponding observed outputs from the system during the historical data (calibration mode, else known as training). While such a dependency is established, it can be next used to predict (or deduce in an effectively way) the unknown future output of the system from the future assumed input values during the application mode (Mitchell 1998).

Within our approach, the NN algorithm is employed to infer the hydrological model errors of the predicted river flows; the error (***e***) is defined here as the difference between observations and simulated flows during validation of the hydrological model (see previous section). This is done by identifying simulated river flow data similar to those from the test data to gain the information on corresponding errors. The underlying assumption here is that the predicted river flows in the future, while using the same hydrological model, will produce similar errors to those observed in the past and therefore it is possible to 'learn' about them from the historical simulated flows and related errors. In view of the assumption that the model error is independent of input and parameter uncertainty, the NN model can be fitted on the error set generated by the optimal hydrological model that has been calibrated.

For the application of the model in prediction mode, initially the deterministic model $\boldsymbol{Q} = S(\boldsymbol{\Theta}, \boldsymbol{X})$ is used which gives a deterministic prediction q_i of the river flow at each time i (step 2 of the simulation procedure in the previous section). Then the space of past data (where the hydrological model was applied in validation mode) is searched for k neighbors (Hastie et al. 2009) nearest to the predicted q_i. The set of neighbors, denoted as $\{N_l(q_i) : l = 1,...,k\}$, form the neighborhood of q_i. The closeness of neighbors is usually expressed by the Euclidean distance (Liu et al. 2004), which for scalar (one-dimensional) data, as in our case, reduces to the absolute distance $|N_l(q_i) - q_i|$. For each one $N_l(q_i)$ of the k nearest neighbors the corresponding errors $e(N_l(q_i))$ in historical simulated river flows are computed. The errors ($E_i := \{e(N_l(q_i)) : l = 1,...,k\}$) infer the distribution of the model error for the predicted river flow q_i.

Next the simulated deterministic prediction q_i is modified by adding a single error value e_i picked at random from the error space E_i, independently for every q_i and assuming all k neighbors equiprobable, thus obtaining a final outcome Q_i:

$$Q_i = q_i + e_i \tag{4}$$

Q_i represents the river flow predicted at the time step i with a simulated random error of the hydrological model. Note that no weighting in error distances is here involved. A vector $\boldsymbol{Q}$ will describe a single realization of the predicted river flow over time according to the simulation procedure presented in the previous section (steps 1-3). This procedure is redone j times (see Fig. 1) where a random error e_i is sampled j-times from the feasible model error space E_i (for each time step). Applying resampling techniques allows therefore obtaining numerous realizations of the error e_i and together with the input and parameter uncertainty (see previous section) covering the prediction limits of the $\boldsymbol{Q}$ expressed in the form of two quantiles of the underlying model prediction distribution $f_Q(\boldsymbol{Q})$ (typically 95%). Note that the E_i is described by a discrete distribution with limited (k) elements. Therefore, because usually $j>>k$, the same model errors can be sampled recursively.

Assumptions and limitations of the NN approach

The proposed approach provides the error which is changing in time and is correlated to the simulated river flow. Note, however, that the simulated errors are conditioned on the magnitude of the river flow alone and no dependency between errors themselves is here explicitly modeled. However, since consecutive outputs q_i of the hydrological model are interdependent, and since, in turn, the error statistics depends on $q_{i,}$,. correlation is implicitly introduced into the error itself, therefore emulating the statistical behavior of the actual error data (see the results presented in Figure 2 below).

One should note that additional driving variables, besides the simulated river flow, may be incorporated into the NN technique to efficiently describe the frequency of occurrence of the model error. For instance, one may consider current or past rainfall, as well as a season, as potential additional information if results obtained with the simulated flow only are not satisfactory. Therefore, testing of the goodness-of-fit of the uncertainty estimation, as performed in the Results section below, is essential to assess the need for additional driving variables.

Although the NN method itself is regarded as nonparametric, it involves a single parameter, the number of nearest neighbors *k*, which has to be specified. This is a sensitive issue. Ideally, *k* should be chosen considering computation time and effect on the statistical characteristics of the model error. Generally, higher values of *k* reduce the effect of noise in searched neighbors. However, too large *k*-values may weaken the dependence of *e* on simulated flow. For a special case when *k* is equal to the size of queried sample, the error will become homoscedastic. Thus, in order to select the proper *k*, a sensitivity analysis as presented in the Results section, rather than a formal parameter fitting procedure, may be required.

It has been noted above that the NN technique proposed here does not involve any weighting factor in searched neighbors. Thus, the choice of NN is made only on the value of the absolute distance between simulated and observed flows. Once located, all *k* NN are equiprobable in sampling. One could think about weighting the contribution of each neighbor according to its distance to the queried q_i. Such weighting, giving a higher weights to NNs located closer can have advantages in regression problems (e.g. Gupta and Mortensen 2009). However, it is not clear if this has a meaning and how it would behave with respect to sampling among different NNs, i.e. the technique considered in this study.

The efficiency of the NN technique is also strongly related to the quality and quantity of the historical data in order to fully (and recursively) cover the river flow variation. Therefore, the method may become less efficient in the case of scarcity or insufficiency of historical data, because it may be difficult to find informative nearest neighbors (Hajebi et al. 2011). Moreover, the method may become slow in deriving a prediction for very long time series (Shrestha and Solomatine 2006). Indeed, many complex hydrological models become slow in evaluation of a big data set. Since for each time step many simulations of the model are to be computed, the computation may become time-consuming (Beygelzimer et al. 2006). Also the NN technique search may become slower if evaluated at very long time series with numerous neighbors. The reason is that all past sample data must be at first scanned at each time step in order to locate the nearest neighbors, for which their corresponding errors are then computed and the resulting error distribution is inferred (for each time step). Not until then random samples from the derived error space can be picked up. Therefore, as we tested our approach on very long time series while in real world applications less data are usually available, the feasibility and usefulness of the method are confirmed. Nevertheless, depending on modeling purposes, a compromise should be sought between the opposite needs to consistently describe the model error and reduce the computational burden.

To accelerate the search of nearest neighbors, we used the kd-tree method, which provides an efficient mechanism for examining only those observations that are closest to the queried, thereby greatly reducing the computation time required to find the closest neighbors (Friedman et al. 1977).

The above proposed technique for sampling from the model error $\boldsymbol{e}$ is based on the assumption that a consistent description of the statistical properties of $\boldsymbol{e}$ can be provided by a sufficiently long sample of model errors themselves that were experienced in validation. Noteworthy, similar assumptions have been recently questioned on the argument that epistemic uncertainty, which affects hydrological models, cannot be represented statistically in view of the fact that disinformative data and epistemic error can lead to short-term non-stationarity in the error statistics that cannot be easily represented by a formal statistical error model with constant parameters (Beven and Westerberg 2011; Beven and Smith 2013, this issue). In our opinion, this line of thought, which implies that epistemic uncertainty is not subject to probabilistic description, may be misleading. Within probability theory, the reason that we use the concept of a random variable is that the quantity of interest is not deterministically known. If a variable is affected by uncertainty, then it is modeled as a random variable, irrespective of the origin of uncertainty. That is, it can be modeled by using stochastics, even if the stochastic dynamics has been imposed due to epistemic uncertainty. This latter may imply the presence of autocorrelation, heteroscedasticity or non-stationarity. Actually, all these are nothing else than stochastic concepts whose definitions are formulated within a stochastic framework. Therefore, invoking these properties to argue about inappropriateness of a stochastic modeling framework is a logical inconsistency, in our view. These properties may increase predictive uncertainty and may underline the need for longer data series for performing statistical inference, but they do not prevent the application of statistical (or data driven) approaches. Therefore, the presence of epistemic uncertainty may affect the results of the proposed approach but does not affect its theoretical validity and does not prevent its practical application.

Case study and experiment setting

Study catchment and data

The presented approach was tested on the catchment of the Secchia River (northern Italy), which is a tributary to the Po River. The contributing area of the analyzed basin to the control section which is conventionally located at Bacchello Bridge (62 km upstream of the confluence in the Po River) is about 1214 km^2 and the length of the stream is of 98 km (Montanari and Koutsoyiannis 2012). The altitude varies from about 30 m to 2121 m above sea level. The mean annual precipitation for this catchment ranges between 700 and more than 2000 mm per year (Montanari 2005).

For the Secchia River basin, historical hourly data on precipitation and temperature both from five rain gauges and river flow at the Bacchello Bridge were available from two years: 1972 and 1973. For the purpose of testing the presented approach, we use synthetic data of 50 years observations generated for the test catchment based on the available historical data as described by Montanari (2005). The synthetic data experiment was adopted mainly for the reason that it enables controlled testing of the influence of epistemic uncertainty, given that we can introduce (and a priori know) it by using different models for the generation of synthetic data and for the method testing. In contrast, in an experiment with real observations it is not possible to know the contribution of epistemic uncertainty because we never know the exact dynamics of the actual (natural) process. Furthermore, the synthetic experiment makes possible the use of an arbitrarily long data set. This is useful in order to be able to test the NN approach with respect to an extended data base and therefore obtaining statistically consistent indications and consistent sensitivity analysis. The same synthetic data set was used by Montanari and Koutsoyiannis (2012) and therefore a comparison of the obtained estimates for simulation uncertainty allows us for consistently quantifying the improvement given by the NN sampling.

The generation of the synthetic data was executed separately for precipitation, temperature and river flow and is briefly presented below.

Generation of the synthetic data

Synthetic precipitation data of 50 years on five rain gauges located within the catchment were generated using the generalized multivariate Neyman-Scott rectangular pulses model (Cowpertwait 1995) which was previously calibrated on the available recorded data from two years (1972 and 1973; (Montanari 2005)). The hourly mean areal rainfall over the basin was calculated as a sum of the rescaled simulated precipitation at five rain gauges using the weights of their contributing polygons with the Thiessen method.

Synthetic data on temperature were obtained for the locations of the available historical data by applying a fractionally differenced autoregressive integrated moving average (FARIMA) model (Montanari et al. 1997). The hourly mean areal temperature over 50 years was computed from the synthetic data through reducing their values to the basin average altitude by adopting a standard temperature gradient of 0.6°C per 100 m of altitude shift.

The generated 50-year synthetic data of precipitation and temperature were next used to obtain synthetic river flow records as outputs from the rainfall-runoff model ADM (Franchini 1996) validated for this catchment. The lumped nine-parameters ADM model was calibrated in previous study against historical data of the Secchia River (1972 and part of 1973) giving satisfying goodness of fit; the Nash-Sutcliffe efficiency (Nash and Sutcliffe 1970) for the validation period (the second part of 1973) was equal to 0.81. Further information on the synthetic data generation for the Secchia River case study and on the ADM model and its calibration can be found in Montanari (2005).These synthetic data for all examined variables are hereinafter referred to as the “actual” data.

Both rainfall and river flow synthetic data were next corrupted in order to account for their uncertainty; the synthetic temperature data were not corrupted due to their limited uncertainty with respect to rainfall and river flow (Montanari and Koutsoyiannis 2012). The hourly mean areal precipitation was corrupted by varying the weights of the contributing rainfall polygons (see also Montanari 2005). This was carried out by randomly picking up at each time step weight values from uniform distributions in the range of $\pm 20\%$ of their uncorrupted values. To retain the cumulative sum of one for all weights, their corrupted values were rescaled at each time step.

The hourly synthetic river flow data generated by the ADM model were corrupted by introducing a multiplier at each time step picked up randomly from a uniform distribution in the range of 0.8-1.2. This allows accounting for the measurement errors in derived river flows.

Hydrological model HyMod

To test our approach we used the commonly applied five parameters conceptual rainfall-runoff model HyMod (Boyle 2000), which was verified before by Montanari (2005) on the same catchment giving satisfactory results.

The HyMod, as a five-parameter model, can be seen as a model of reduced complexity in comparison to the nine-parameter ADM model. Therefore, it will presumably not perfectly reproduce the synthetic river flow data generated by the ADM model. As a result, the output of the HyMod will be contaminated by error due to its simplified model structure with respect to the ADM model. This can be regarded as epistemic uncertainty, according to the common perception of what the latter is, given that the original data are not natural but synthetic and were produced by a different model, which is perfectly accurate with respect to these data (because it produced them). The simplified HyMod model does not perfectly represent the original (ADM) dynamics, thus giving rise to imperfections of the dynamical description.

The inputs into the HyMod are mean areal precipitation and evapotranspiration. Evapotranspiration is here considered using the radiation method as proposed by Doorenbos et al. (1984).

Simulation procedure and prediction limits generation

The proposed approach was tested on the synthetic data derived as described above. The available 50 years dataset was split into three periods in order to: (1) calibrate the HyMod (calibration period, years: 1-30), (2) infer the error model of HyMod structural deficits in validation mode (error identification period, years: 31-40) and (3) fully validate the approach (testing period, years: 41-50).

The calibration of the HyMod was carried out by using a classical approach. Specifically, we used the DREAM algorithm (Vrugt and Robinson 2007) to minimize the sum of the squared residuals in the simulation of the synthetic data over the first 30 years (years: 1-30). The Nash-Sutcliffe efficiency was 0.93 for the validation when using years 31-50. In principle, more rigorous likelihood functions could be used for model calibration instead of least squares (see, for instance, the solutions proposed by Schoups and Vrugt (2010) and Pianosi and Raso (2012)). However, they would imply the introduction of additional parameters which we preferred to avoid in order to carry out a consistent comparison with the results presented by Montanari and Koutsoyiannis (2012).

It is well known that the sum of the squared residuals is an approximation of the Gaussian likelihood function and therefore one may note that we indirectly used a specific likelihood for parameter calibration, which we dismissed when estimating the uncertainty of the model predictions. Therefore, it is relevant to point out that such a procedure is not inconsistent. Assumptions that are acceptable for parameter estimation may be no more justified when estimating the global uncertainty, because of the different impact that the same assumption may have on different procedures of statistical inference. In this paper it is our intention not to use a formal likelihood for estimating prediction uncertainty.

By sampling randomly from the posterior parameters distribution $f_{\Theta}(\boldsymbol{\Theta})$ we obtained 6000 realizations of the parameter set, which we then used over the validation period to test our approach (years: 41-50). Sampling a parameter vector instead of single parameters independently allows retaining the mutual dependencies between all parameters.

The years 31-40 were used to infer the space of the model structure errors E_i. To fully explore the error space, we conducted the resampling technique at every time step of the predicted river flow (testing period) as described above in section "The Nearest Neighbor approach".

The latter period (years 41-50) was used to fully validate the proposed approach. For this period 1000 simulations were executed, each one by picking up randomly a set of HyMod's parameters from the posterior parameters space $f_{\Theta}(\boldsymbol{\Theta})$ and one realization of the corrupted rainfall as described above. Therefore, input uncertainty has been assumed to be known, in order to be able to focus on the efficiency of the NN technique for sampling from the model error. Rainfall and the corresponding evaporation were passed through the HyMod. Outputs of the hydrological model were next modified by adding one realization of the predicted river flow errors E_i derived at each time step using the NN method as described earlier.

From the results of this Monte Carlo (MC) simulation with 1000 repetitions for the 10-year testing period (years: 41-50), the corresponding prediction limits were computed as the two quantiles i.e. 97.5% and 2.5%, providing jointly the 95% prediction limits band. A warm-up period at the beginning of the testing period, with the length of three months, was excluded from the further analysis.

Assessment of the hydrological predictions reliability

The reliability of model predictions was assessed under the Nash-Sutcliffe efficiency and root mean square error (RMSE) for the best prediction of the simulations (mode) for two cases: before and after modifying the outputs from the HyMod with a random error using the NN approach.

For the practical use of the NN approach, we carried out a sensitivity analysis of the model response for a few different values of the k parameter. Specifically, we performed 1000 MC simulations for six different values of k, namely k = 5, 10, 20, 30, 50, 100, each time computing the 95% prediction limits and their observation coverage, as well as the percentage of values lying within the prediction limits, which theoretically should be 95%.

The adequacy of the derived prediction limits (for the fixed *k*) was assessed by the coverage probability method as proposed by Laio and Tamea (2007). This method relies on the assumption that the probability density function of empirical distribution quantiles of a predicted variable is uniform (U(0,1)). This means that the variable should be overestimated and underestimated with the same probability. If it is so, the prediction limits should be symmetrically spread along the central value (50% quantile). The coverage probability can be practically assessed from the Coverage Probability Plot (CPP) that presents the theoretical against the computed quantiles. The deviation of plotted points from the bisector line (1:1) allows locating areas where predictions are systematically overestimated or underestimated. Ideally, the empirical points should coincide with the bisector line indicating that the model prediction limits are uniform and consistent with the theoretical 95% data coverage in the entire range of river flows and over the entire period.

Results

Diagnosis of simulated errors

The diagnostic plot of the residuals simulated with the NN technique is provided in Fig. 2 and compared to the residuals of the calibrated hydrological model (HyMod). A plot of residuals versus simulated values (right panel in Fig. 2) shows that points are randomly scattered around the value of 0. However, a correlation between the magnitude of residuals and the values of simulated flows is observed. This is in agreement with the assumption underlying the proposed NN technique, which relates model errors to the simulated (observed) river flows. The autocorrelation of residuals is presented on the left panel in Fig. 2. Although the non-parametric NN technique does not explicitly account for the error autocorrelation, this is preserved in simulated errors. The autocorrelation of simulated residuals (Fig. 2, left bottom) is noticeable and in agreement with residuals of the calibrated model (Fig. 2, left top). This proves that the NN technique effectively simulates hydrological model errors by relating errors to the simulated flows and thus indirectly accounting for their correlation (present in simulated flows).

Model prediction efficiency

Correcting simulated river flow via the NN method slightly improved the model predictions (see Fig. 3). The corresponding Nash-Sutcliffe efficiency was found to be 0.83 against 0.82 (without modifying river flows) for the best prediction (mode), while RMSE was reduced by about 27% (from 14.4 to 10.4 $m^3 s^{-1}$). The improvements are especially visible for the peaks, which are of the highest concern in flood predictions and preventions. Note, however, that the main objective of the method is not the improvement of the prediction accuracy, but the conversion of a deterministic prediction into probabilistic one by providing confidence limits for the predicted variable.

Sensitivity to the *k* value of the NN approach

The sensitivity analysis proved that the *k* value, among the reasonable range between 5 and 100 considered, has very little influence on the simulated river flows and the corresponding prediction limits. The prediction limits derived with different values of *k* varied by less than 0.1% of the total observation coverage. This can be explained, first, by the goodness of the model fit to the “actual” data and therefore similar model errors (deviations between simulated and “actual” river flows) over the entire river flows. And second, in the case of a training data set with a sufficient length (covering fully the river flow variation), the differences in model errors estimated by assigned nearest neighbors for different *k* may be considered statistically indifferent. The explanation is that when k is much smaller than the calibration series length, it is always possible to find within the calibration set at least k well-fitting neighbors. Therefore, to minimize the computation effort, we limited our analysis to the $k = 10$ case. The sensitivity of the predictions to the k-value may, however, need to increase in a situation of limited data for the NN search or when the model does not explain satisfactorily the behavior of a catchment.

Percentage coverage of the prediction limits

The 95% prediction limits computed by MC simulation are shown in Fig. 4. These cover 97.9% of all observations for the testing period against the theoretical value of 95%. That means that the intervals between prediction limits are slightly overestimated. In particular, 0.5% and 1.6% of the “actual” points lie above and below the prediction limits respectively against the theoretical values of 2.5% for each of the intervals. This is a satisfying achievement and a noticeable improvement in comparison to the recent study of Montanari and Koutsoyiannis (2012) on the same catchment but with a different method of modifying simulated river flow (the Meta-Gaussian approach by Montanari and Brath (2004)), where 5.4% and 4.3% of the “actual” river flows fell, respectively, out of the upper and lower 95% prediction limits.

Verification of the coverage probability

The coverage probability of the predictions was evaluated for 1000 repetitions of the simulated river flows over 10 years of the validation period for $k = 10$. Figure 5 presents the resulting coverage probability plot (CPP) for the case study (gray line). As can be seen from the figure, the confrontation of the computed quantiles with the theoretical ones indicates satisfactory predictions of the variable; the points lie along and close to the bisector line (1:1), especially for lower quantiles, whereas, a slight overestimation of the predicted river flow for higher quantiles is observed. Thus, the derived prediction limits may be considered as reliable in flood forecasting. This is again a significant improvement in comparison with the previous study of Montanari and Koutsoyiannis (2012), where the predictions were visibly more underestimated for all quantiles (compare to the black line in Fig. 5).

Concluding remarks

An original procedure for sampling outcomes from the error population of hydrological models was incorporated within the modeling framework proposed by Montanari and Koutsoyiannis (2012). Specifically, the model error is assumed to represent all sources of uncertainty (epistemic or induced by inherent variability) other than input and parameter uncertainty. Therefore, sampling from the model error allows a reliable reconstruction of the probability distribution of the model output, provided the complex statistical properties of the error itself are preserved. The idea explored relies on the use of a nearest neighbor resampling procedure from realizations of the hydrological model errors in a past period but not used for the model calibration itself.

The results and the statistical assessment that have been performed to check the reliability of the estimated confidence bands for model simulation prove that the proposed procedure leads to an efficient uncertainty assessment. In fact, the above statistical tests, namely the coverage probability plot and the computation of observed data lying between the confidence bands, indicate that a considerable improvement was reached with respect to the results obtained by Montanari and Koutsoyiannis (2012), who instead used the Meta-Gaussian error model to extract random outcomes from the error population.

The results confirm that error resampling techniques may be an interesting option to account for prediction uncertainty thereby avoiding a formal statistical characterization of the model error, when it is difficult to parameterize. The proposed approach presents the limitation of requiring a sufficiently long enough data records, first, for the hydrological model calibration and second, for the error characterization using the nearest neighbor technique. In order to provide reliable estimations, the same data set cannot be used twice.

The proposed approach relies much on data. In particular, to obtain a proper characterization of the distributional properties of the model error through resampling techniques a fairly extended data base of previous simulation errors for the considered (and calibrated) hydrological model is needed. Herein the difficulties related to the availability of a consistent data base were not considered, because the testing of the proposed approach was intentionally based on synthetic data. Ongoing research is focusing on real world applications for catchments where historical data are available for an extended observation period.

Acknowledgements

The work presented herein was carried out within the COST Action ES0901 titled "European Procedures for Flood Frequency Estimation (FloodFreq)" and the Panta Rhei Research Initiative of the International Association of Hydrological Sciences (IAHS). A financial support by EU for A.E.S. is gratefully acknowledged. A.E.S. would also like to acknowledge Prof. Kazimierz Banasik (WULS-SGGW, Poland) for his input in the overall execution of this project. We thank the eponymous reviewer Keith Beven and an anonymous reviewer, as well as the Guest Editor Sivakumar Bellie, for providing detailed and useful comments; although we disagree with a few of the comments, overall they helped us to improve the paper substantially.

References

Ajami, N. K., Duan, Q., and Sorooshian, S. (2007). "An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction.” *Water Resources Research*, 43(W01403), 19 pp.

Beven, K.J. (2006). "A manifesto for the equifinality thesis", *Journal of Hydrology*, 320, 18-36.

Beven, K.J. (2010). "Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models", *Hydrological Processes*, 24, 1537–1547).

Beven, K.J. (2012). "Causal models as multiple working hypotheses about environmental processes", *Comptes Rendus Geoscience*, 344, Pages 77-88.

Beven, K. J., and Binley, A. M. (1992). “The future of distributed models: model calibration and uncertainty prediction.” *Hydrological Processes*, 6, 279–298.

Beven K. J. and Smith P. (2013), *Journal of Hydrologic Engineering*, this issue. Reference to be added once known.

Beven, K. J. and Westerberg, I. K. (2011). “On red herrings and real herrings: disinformation and information in hydrological inference”, *Hydrological Processes*, 25, 1676-1680.

Beven, K., Smith P., Westerberg I., and Freer J. (2012). "Comment on ‘‘Pursuing the method of multiple working hypotheses for hydrological modeling’’ by P. Clark et al.", Water Resources Research, 48, W11801, doi:10.1029/2012WR012282.

Beygelzimer, A., Kakade, S., and Langford, J. (2006). “Cover trees for nearest neighbor.” *ACM Proc. 23rd international conference on Machine learning*, 97–104.

Blöschl, G., Grayson, R.B. and Sivapalan, M. (1995) “On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling." *Hydrological Processes*, 9, 313–330.

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

Boyle, D. P. (2000). “Multicriteria calibration of hydrological models.” *Ph.D. Dissertation,* Dep. of Hydrol. and Water Resour., Univ. of Ariz.

Brath, A., Montanari, A., and Toth, E. (2002). “Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models.” *Hydrology and Earth System Sciences Discussions*, 6(4), 627–639.

Bulygina, N., and H. Gupta *(*2009*),* Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resources Research*, 45, W00B13, doi:10.1029/2007WR006749.

Bulygina, N., and Gupta, H. (2010). “How Bayesian data assimilation can be used to estimate the mathematical structure of a model.” *Stochastic Environmental Research and Risk Assessment*, 24(6), 925–937.

Bulygina, N., and H. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resources Research*, 47, W05514, doi:10.1029/2010WR009614.

Clark, M. P., and Slater, A. G. (2006). “Probabilistic quantitative precipitation estimation in complex terrain.” *Journal of Hydrometeorology*, 7(1), 3–22.

Clark, M. P., Slater A. G., Rupp D. E., Woods R. A., Vrugt J. A., Gupta H. V., Wagener T., and Hay L. E. (2008). "Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models". *Water Resources Research*, 44, W00B02, doi:10.1029/2007WR006735.

Clark, M. P., Kavetski D., and Fenicia F. (2011). "Pursuing the method of multiple working hypotheses for hydrological modeling." *Water Resources Research*, 47, W09301, doi:10.1029/2010WR009827.

Cowpertwait, P. S. P. (1995). “A Generalized Spatial-Temporal Model of Rainfall Based on a Clustered Point Process.” *Proceedings of the Royal Society A*, 450, 163– 175.

Cunha, L.K., Mandapaka, P.V., Krajewski, W.F., Mantilla, R., Bradley, A. A., (2012). "Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model." *Water Resources Research*, 48 doi: 10.1029/2012WR012138.

Di Baldassarre, G., and Montanari, A. (2009). “Uncertainty in river discharge observations: a quantitative analysis.” *Hydrology and Earth System Sciences*, 13(6), 913–921.

Doorenbos, J., Pruitt, W. O., Aboukhaled, A., Damagnez, J., Dastane, N. G., Van der Berg, C., Rijtema, P. E., Ashford, O. M., and Frere, M. (1984). “Guidelines for predicting crop water requirements.” *FAO Irrigation and Drainage Paper.*

Ebtehaj, M., Moradkhani, H., and Gupta, H. V. (2010). “Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling.” *Water Resources Research*, 46(W07515), 14 pp.

Franchini, M. (1996). “Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models.” *Hydrological Sciences Journal-journal Des Sciences Hydrologiques*, 41(1), 21–39.

Friedman, J. H., Bentley, J. L., and Finkel, R. A. (1977). “An algorithm for finding best matches in logarithmic expected time.” *ACM Transactions on Mathematical Software (TOMS)*, 3(3), 209–226.

Gupta M.R. and Mortensen W.H. (2009). "Weighted Nearest Neighbor Classifiers and First-Order Error," *Proc. Int'l Conf. Frontiers of Interface between Statistics and Science*, pp. 1-21.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M. (2012). “Towards a comprehensive assessment of model structural adequacy.” *Water Resources Research*, 48(W08301), 16 pp.

Hajebi, K., Abbasi-Yadkori, Y., Shahbazi, H., and Zhang, H. (2011). “Fast approximate nearest-neighbor search with k-nearest neighbor graph.” *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, 1312–1317.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). "Elements of Statistical Learning: data mining, inference, and prediction". *Springler Serie in Statistics*, Springer.

He, X., Refsgaard, J. C., Sonnenborg, T. O., Vejen, F. and Jensen K. H. (2011). "Statistical analysis of the impact of radar rainfall uncertainties on water resources modeling". *Water Resources Research,* 47*,* W09526, doi:10.1029/2011WR010670.

Karlsson, M., and Yakowitz, S. (1987a). “Nearest-neighbor methods for nonparametric rainfall-runoff forecasting.” *Water Resources Research*, 23(7), 1300–1308.

Karlsson, M., and Yakowitz, S. (1987b). “Rainfall-runoff forecasting methods, old and new.” *Stochastic Hydrology and Hydraulics*, 1(4), 303–318.

Koutsoyiannis, D., (2013). "Hydrology and Change", *Hydrological Sciences Journal*, 58(6), 1177-1197, doi:10.1080/02626667.2013.804626.

Koutsoyiannis, D., Yao, H., and Georgakakos, A. (2008). “Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods.” *Hydrological Sciences Journal*, 53(1), 142–164.

Krueger, T., Freer, J., Quinton, J. N., Macleod, C. J. A., Bilotta, G. S., Brazier, R. E., Butler, P., and Haygarth, P. M. (2010). “Ensemble evaluation of hydrological model hypotheses.” *Water Resources Research*, 46(W07516), 17 pp.

Krzysztofowicz, R. (2002). “Bayesian system for probabilistic river stage forecasting.” *Journal of Hydrology*, 268(1-4), 16–40.

Laio, F., and Tamea, S. (2007). “Verification tools for probabilistic forecasts of continuous hydrological variables.” *Hydrology and Earth System Sciences Discussions*, 11(4), 1267–1277.

Legleiter, C. J., Kyriakidis, P. C., McDonald, R. R., and Nelson J. M. (2011). "Effects of uncertain topographic input data on two-dimensional flow modeling in a gravel-bed river". *Water Resources Researc*h, 47, W03518, doi:10.1029/2010WR009618.

Liu, T., Moore, A. W., Gray, A., and Yang, K. (2004). “An investigation of practical approximate nearest neighbor algorithms.” *Advances in neural information processing systems*, 17, 825–832.

Liu, Y., Freer, J., Beven, K., and Matgen, P. (2009). “Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error.” *Journal of Hydrology*, 367(1-2), 93–103.

Mantovan, P., and Todini, E. (2006). “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology.” *Journal of Hydrology*, 330(1-2), 368–381.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., and Clark, M. (2010). “Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions.” *Hydrological Processes*, 1270–1284.

McMillan, H., Jackson, B., Clark, Martyn, Kavetski, D., and Woods, R. (2011). “Rainfall Uncertainty in Hydrological Modelling: An Evaluation of Multiplicative Error Models | Mendeley.” *Journal of Hydrology*, 400(1-2), 83–94.

Mendoza, P. A., McPhee, J., Vargas, X. (2012). "Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment", *Water Resources Research*, 48, W09532, doi:10.1029/2011WR011089.

Mitchell, T. M. (1998). Machine learning. *New York: McGraw-Hill*.

Montanari, A. (2005). “Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations.” *Water Resources Research*, 41(8).

Montanari, A. (2011). “Uncertainty of Hydrological Predictions.” Peter Wilderer (ed.) Treatise on Water Science, *Oxford Academic Press*, 459–478.

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth,Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D., Srinivasan, V., Harman, C., Thompson, S., Rogger,Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V. (2013). "“Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022", *Hydrological Sciences Journal*, 58 (6), doi: 10.1080/02626667.2013.809088.

Montanari, A., and Brath, A. (2004). “A stochastic approach for assessing the uncertainty of rainfall-runoff simulations.” *Water Resources Research*, 40(1), W01106.

Montanari, A. and Di Baldassarre, G. (2013). "Corrigendum to data errors and hydrological modelling: The role of model structure to propagate observation uncertainty". *Advances in Water Resources*, 51, 498-505, doi: 10.1016/j.advwatres.2012.09.007

Montanari, A., and Grossi, G. (2008). “Estimating the uncertainty of hydrological forecasts: A statistical approach.” *Water Resources Research*, 44(12), W00B08.

Montanari, A., and Koutsoyiannis D. (2012), “A blueprint for process-based modeling of uncertain hydrological systems”. *Water Resource Research*, 48, W09555, doi:10.1029/2011WR011412.

Montanari, A., Rosso, R., and Taqqu, M. S. (1997). “Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation.” *Water Resources Research*, 33(5), 1035–1044.

Moretti G., and Montanari, A. (2008). "Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model", *Hydrology and Earth System Sciences*, 12, 1141-1152.

Nash, J. E., and Sutcliffe, J. V. (1970). “River flow forecasting through conceptual models. Part-I: a discussion of principles.” *Journal of Hydrology*, 10(3), 282–290.

Neuman, S. P. (2003). “Maximum likelihood Bayesian averaging of uncertain model predictions.” *Stochastic Environmental Research and Risk Assessment*, 17(5), 291–305.

Pianosi, F., and L. Raso (2012), Dynamic modeling of predictive uncertainty by regression on absolute errors, *Water Resources Research*, 48, W03516, doi:10.1029/2011WR010603.

Renard, B., Kavetski, D., Leblois, M., Thyer, M., Kuczera, G., and Franks, S. W. (2011). “Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation.” *Water Resources Research*, 47(W11516), 21 pp.

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., Franks, S.W. (2011). "Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resources Reswearch*, 47, W11516, doi:10.1029/2011WR010643.

Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and nonGaussian errors, *Water Resources Research,* 46, W10531, doi:10.1029/2009WR008933.

Shrestha, D. L., Kayastha, N., and Solomatine, D. P. (2009). “A novel approach to parameter uncertainty analysis of hydrological models using neural networks.” *Hydrology and Earth System Sciences*, 13(7), 1235.

Shrestha, D. L., and Solomatine, D. P. (2006). “Machine learning approaches for estimation of prediction interval for the model output.” *Neural Networks*, 19(2), 225–235.

Skøien, J. O., Blöschl, G., and Western, A. W. (2003). "Characteristic space scales and timescales in hydrology." *Water Resources Research*, 39, article number 1304, doi:10.1029/2002wr001736.

Sikorska, A. E. (2012). “Interactive comment on ‘Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models’ by AE Sikorska et al.” *Hydrology and Earth System Sciences Discussions*, 8, C6284–C6310.

Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J. (2012). “Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models.” *Hydrology and Earth System Sciences*, 16, 1221–1236.

Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J. (2013). "Considering rating curve uncertainty in water level predictions". *Hydrology and Earth System Sciences Discussions*, 10, 2955-2986, doi:10.5194/hessd-10-2955-2013, 2013.

Srikanthan, R., Kuczera, G., and Thyer, M. (2009). “‘Calibrate It Twice’: A Simple Resampling Method for Incorporating Parameter Uncertainty in Stochastic Data Generation.” *Proceedings of H2O 09: 32nd Hydrology and Water Resources*

Symposium, Newcastle: Adapting to Change., Barton, A.C.T.: Engineers Australia, 2009, 1028–1039.

Sun, S., and Bertrand-Krajewski, J-L. (2013). "Separately accounting for uncertainties in rainfall and runoff: Calibration of event based conceptual hydrological models in small urban catchments using Bayesian method", *Water Resources Research*, 49, http://dx.doi.org/10.1002/wrcr.20444, in press (available on-line).

Tada, T., and Beven, K. J. (2012). “Hydrological model calibration using a short period of observations.” *Hydrological Processes*, 26(6), 883–892.

Toth, E., Montanari, A., and Brath, A. (1999). “Real-time flood forecasting via combined use of conceptual and stochastic models.” *Physics and Chemistry of the Earth (B)*, 24(7), 793–798.

Vrugt, J. A., and Robinson, B. A. (2007). “Improved evolutionary optimization from genetically adaptive multimethod search.” *Proceedings of the National Academy of Sciences*, 104(3), 708–711.

Yang, J., Reichert, P., Abbaspour, K. C., and Yang, H. (2007). “Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference.” *Journal of Hydrology*, 340(3-4), 167–182.

Winsemius, H.C., Savenije, H.H.G., Gerrits, A.M.J., Zapreeva, E.A., and Klees, R. (2006). "Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability." *Hydrology and Earth System Sciences*, 10: 339--352.

Winsemius, H.C., Savenije, H.H.G., and Bastiaanssen, W.G.M. (2008). "Constraining model parameters on remotely sensed evaporation: Justification for distribution in ungauged basins?" *Hydrology and Earth System Sciences*, 12: 1403--1413.

Winsemius, H.C., Schaefli, B., Montanari, A., Savenije, H.H.G. (2009). "On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information". *Water Resources Research*, 45: doi: 10.1029/2009WR007706,2009.

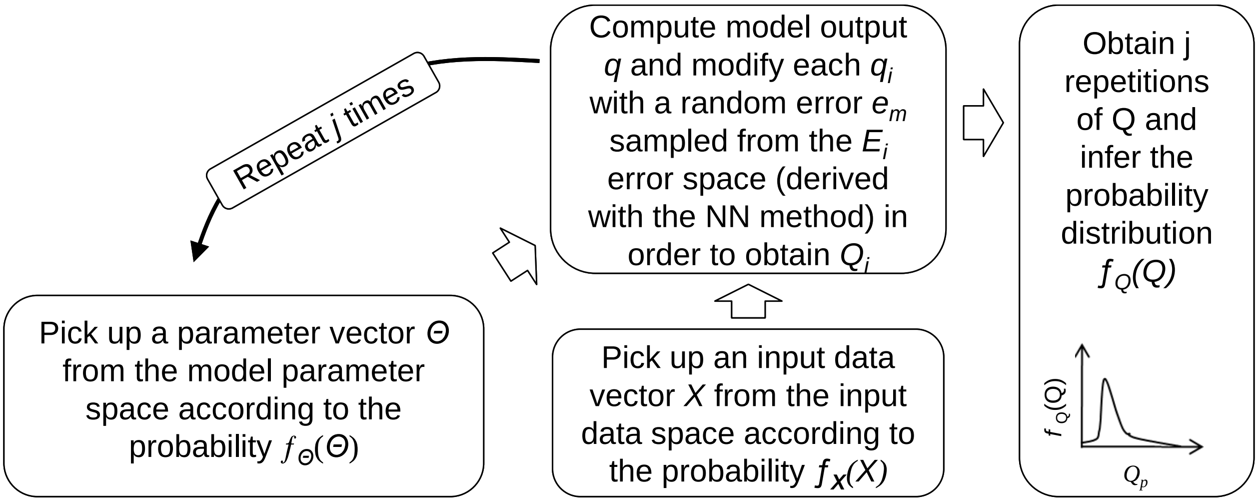

Fig. 1. Flowchart of the Monte Carlo simulation for estimating prediction limits

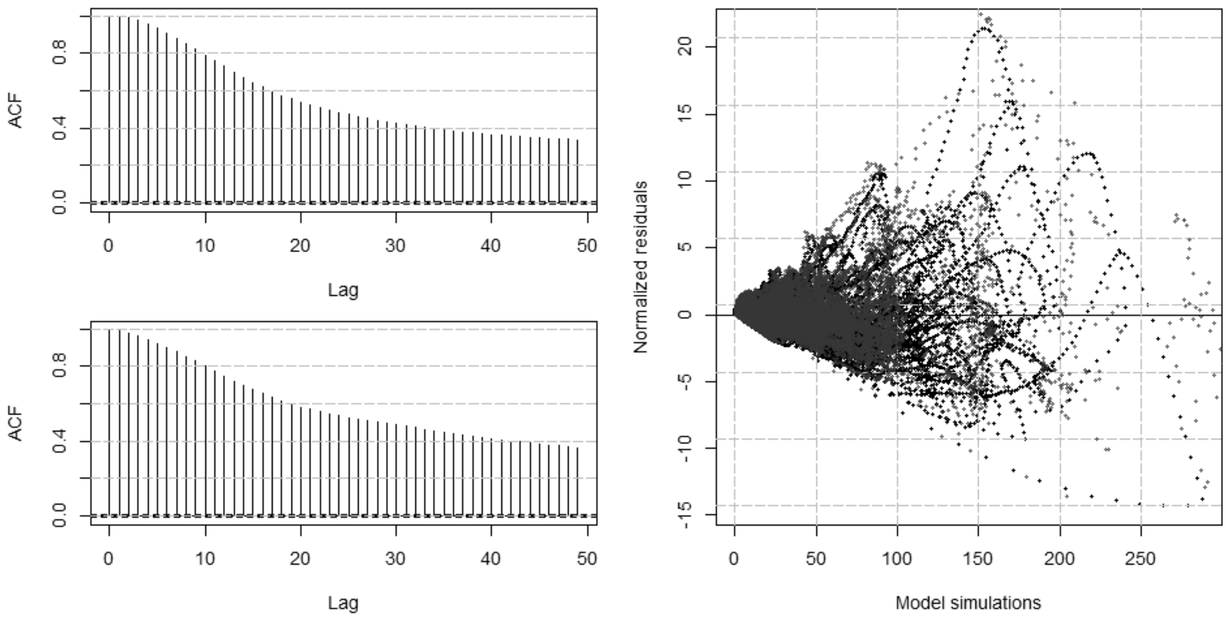

Fig. 2. Diagnostic plot of the residuals; left panel: autocorrelation function (ACF) of residuals of calibrated model (top) and simulated residuals with the NN-method (bottom); right panel: black points - normalized residuals vs. model simulations (calibrated), gray points – simulated residuals vs. model simulations with simulated residuals.

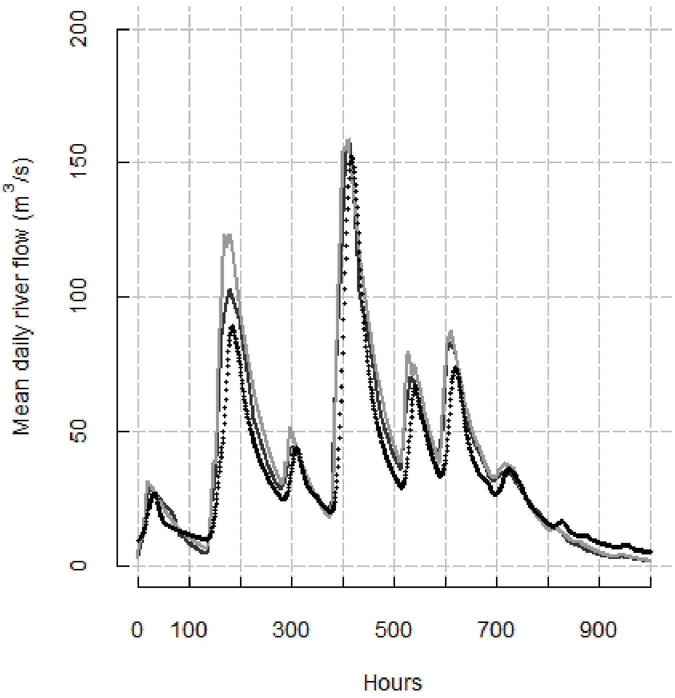

Fig. 3. "Actual" and predicted river flow for the Secchia River over 1000 hours out of the 10 years (41-50) testing period. The black dotted line corresponds to the "actual" river flows, the black continuous line to the best prediction (mode) of river flow modified via NN and the continuous gray line to the best prediction without any modification of the river flow.

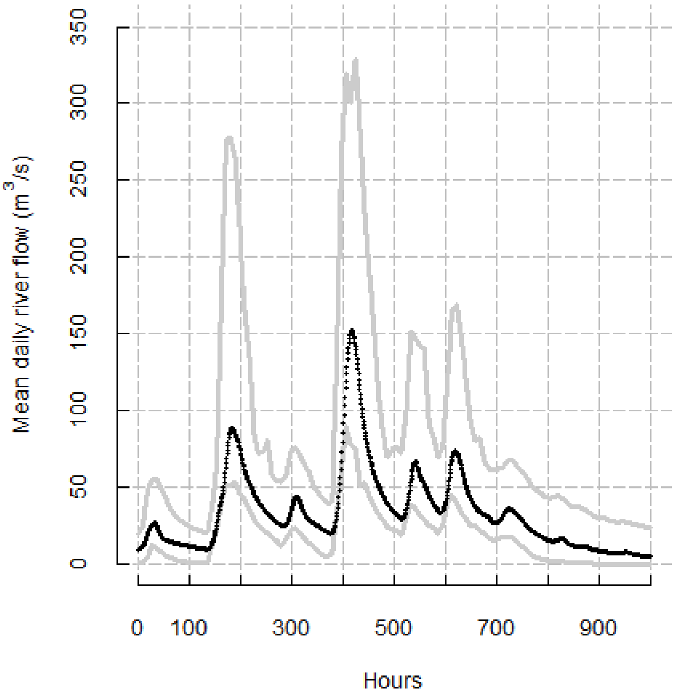

Fig. 4. 95% - prediction limits for the Secchia River during 1000 hours out of the 10 years (41-50) testing period. The black dotted line corresponds to the “actual” river flows and the gray continuous lines to the 95% prediction limits.

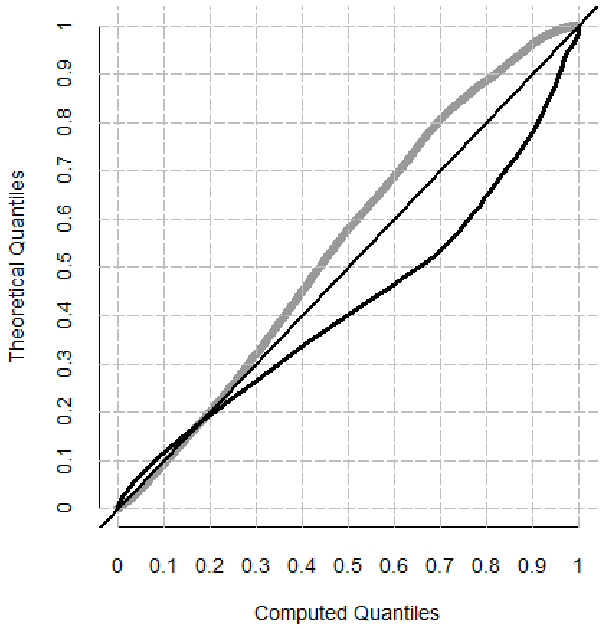

Fig. 5. Coverage Probability Plot for the case study of the Secchia River over the testing period of 10 years (41-50) for predicted river flows; gray line –with simulated errors whilst using NN approach, black line –with simulated errors while using Meta-Gaussian approach (reproduced from Montanari and Koutsoyiannis, 2012).