

1
2
3
4
5
6
7
8
9
10
11
12
13

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

Anna E. Sikorska¹, Alberto Montanari ² and Demetris Koutsoyiannis³

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 Polytechniou 5, GR 157 80, Zographou, Greece. E-mail: dk@itia.ntua.gr

14 **Abstract**

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

30

31

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

34

35

36

37 **Introduction**

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

47 The problem of uncertainty assessment is in principle reducible to estimating and
48 integrating the main sources of uncertainty through the modeling chain. The literature
49 proposed several contributions on the estimation of input uncertainty (Clark and Slater 2006;
50 Cunha et al. 2012; He et al. 2011; Legleiter et al. 2011; McMillan et al. 2011; Montanari and
51 Di Baldassarre 2013; Renard et al. 2011; Sikorska et al. 2012; Sun and Bertrand-Krajewski
52 2013), parameter uncertainty (Ebtehaj et al. 2010; Srikanthan et al. 2009; Vrugt and Robinson
53 2007), measurement errors/output uncertainty (Di Baldassarre and Montanari 2009;
54 McMillan et al. 2010; Sikorska et al. 2013) and uncertainty in the model structure
55 (Krzysztofowicz 2002; Montanari and Brath 2004; Montanari and Grossi 2008), but the
56 related problems are far from being solved.

57 Actually, the estimation of input and parameter probability distributions for
58 hydrological models through the above mentioned methods is often affected by the presence
59 of model structural errors which in turn are themselves related to data errors and parameter
60 errors. Indeed, individual contributions to the global uncertainty cannot be quantified
61 independently, unless (1) one introduces assumptions about the nature of the individual error

62 components or (2) observations are available that allow estimating each source of error
63 independently (Renard et al. 2011). For instance, the latter case was examined by Winsemius
64 et al. (2006; 2008; 2009) who used gravity and evaporation measurements to constrain
65 parameter estimation for a rainfall-runoff model. In the former case, the assumptions that are
66 introduced to estimate each uncertainty source separately can affect the reliability of
67 estimations (Beven 2006; 2010; 2012). A comprehensive review of uncertainty assessment
68 techniques and their underlying assumptions has been recently presented by Montanari
69 (2011), while the issue of introducing assumptions in uncertainty assessment in hydrology
70 has been further discussed by Montanari and Koutsoyiannis (2012).

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

82 As for the integration of the different uncertainty sources, several methods adopt a
83 numerical procedure. According to that, the probability distribution of the model outcome is
84 estimated by producing several outputs, which are obtained by randomly sampling the
85 feasible spaces of input data, parameters and model structure errors usually merged together

86 with measurement errors (Ajami et al. 2007; Krueger et al. 2010; Liu et al. 2009; Neuman
87 2003).

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

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

109 A generalized approach (a blueprint) to carry out the uncertainty assessment for model
110 predictions has been recently proposed by Montanari and Koutsoyiannis (2012). By admitting

111 that uncertainty in hydrological models mainly originates from model structural inadequacy
112 (which descends from limited knowledge and therefore is an epistemic form of uncertainty),
113 uncertainty in the observed data (input and output observations) and inherent randomness
114 (e.g. due to sensitive dependence on initial conditions), the basic assumption in the above
115 blueprint lies in the recognition that randomness is an intrinsic property of hydrological
116 processes. In this respect, randomness could be thought of as indeterminacy or inherent
117 inability to describe the future evolution of hydrological processes deterministically. The
118 notion of indeterminacy is used here to underline our belief that a perfect reproduction of
119 hydrological processes at scales of practical interest will never be possible and therefore the
120 model error is also the result of an intrinsic property rather than just model inadequacy.

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

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

135 interpreting the heteroscedasticity of the error itself. Thus, the statistical characterization of
136 the model error still remains a problem (Montanari and Koutsoyiannis 2012).

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

149 Application of the NN method in hydrology is not new. In fact, the NN method has
150 found already implementations in a wide range of real-world settings as pattern recognition,
151 machine learning and database querying (Liu et al. 2004; Shrestha et al. 2009; Shrestha and
152 Solomatine 2006) and for searching a model space (Beven and Binley 1992). Karlsson and
153 Yakowitz (1987a) have demonstrated the usefulness of the NN method to large-sample time
154 series problems. Due to its intuitiveness, simplicity (non-parametric property) and the sound
155 theoretical basis, it has been made also attractive to forecasters in the hydrologic field for
156 time series predictions (Brath et al. 2002; Karlsson and Yakowitz 1987b; Koutsoyiannis et al.
157 2008; Toth et al. 1999).

158 The approach proposed in this paper is tested by using 50 years of synthetic data
159 referred to a river basin in Italy (Secchia River). The obtained results show that NN is a very

160 efficient solution for solving the problem of characterizing the model error in hydrological
161 predictions for long time series when sufficient historical data are available.

162

163 **Theoretical setting for uncertainty assessment**

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

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

$$172 \quad \mathbf{Q} = S(\boldsymbol{\theta}, \mathbf{X}) \tag{1}$$

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

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

$$182 \quad f_{\mathbf{Q}}(\mathbf{Q}) = K_s f_{\boldsymbol{\theta}, \mathbf{X}}(\boldsymbol{\theta}, \mathbf{X}) \tag{2}$$

183 where f indicates a probability density function and K_S is a stochastic operator which depends
 184 on (but is different from) the deterministic model S . In detail, Montanari and Koutsoyiannis
 185 (2012) proved that the deterministic Eq. (2) can be converted to the stochastic form

$$186 \quad f_Q(\mathbf{Q}) = \int_{\Theta} \int_X f_e(\mathbf{Q} - S(\Theta, \mathbf{X}) | \Theta, \mathbf{X}) f_{\Theta}(\Theta) f_X(\mathbf{X}) d\Theta d\mathbf{X} \quad (3)$$

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

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

205 We may note that the proposed simulation procedure is similar to GLUE (Beven and
 206 Binley 1992; Liu et al. 2009) with the exception that GLUE rejects non-behavioral

207 simulations by usually adopting a likelihood measure. Moreover, one may note that in many
208 applications GLUE was used without including the random contribution of the model error in
209 the formation of the output uncertainty.

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

215

216 **Sampling from the probability distribution of the model error**

217 **The Nearest Neighbor approach**

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

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

241 For the application of the model in prediction mode, initially the deterministic model
242 $Q = S(\theta, X)$ is used which gives a deterministic prediction q_i of the river flow at each time i
243 (step 2 of the simulation procedure in the previous section). Then the space of past data
244 (where the hydrological model was applied in validation mode) is searched for k neighbors
245 (Hastie et al. 2009) nearest to the predicted q_i . The set of neighbors, denoted as
246 $\{N_l(q_i) : l = 1, \dots, k\}$, form the neighborhood of q_i . The closeness of neighbors is usually
247 expressed by the Euclidean distance (Liu et al. 2004), which for scalar (one-dimensional)
248 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
249 nearest neighbors the corresponding errors $e(N_l(q_i))$ in historical simulated river flows are
250 computed. The errors ($E_i := \{e(N_l(q_i)) : l = 1, \dots, k\}$) infer the distribution of the model error
251 for the predicted river flow q_i .

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

255 $Q_i = q_i + e_i$ (4)

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

267 **Assumptions and limitations of the NN approach**

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

275 One should note that additional driving variables, besides the simulated river flow, may
276 be incorporated into the NN technique to efficiently describe the frequency of occurrence of
277 the model error. For instance, one may consider current or past rainfall, as well as a season,
278 as potential additional information if results obtained with the simulated flow only are not

279 satisfactory. Therefore, testing of the goodness-of-fit of the uncertainty estimation, as
280 performed in the Results section below, is essential to assess the need for additional driving
281 variables.

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

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

299 The efficiency of the NN technique is also strongly related to the quality and quantity
300 of the historical data in order to fully (and recursively) cover the river flow variation.
301 Therefore, the method may become less efficient in the case of scarcity or insufficiency of
302 historical data, because it may be difficult to find informative nearest neighbors (Hajebi et al.
303 2011). Moreover, the method may become slow in deriving a prediction for very long time

304 series (Shrestha and Solomatine 2006). Indeed, many complex hydrological models become
305 slow in evaluation of a big data set. Since for each time step many simulations of the model
306 are to be computed, the computation may become time-consuming (Beygelzimer et al. 2006).
307 Also the NN technique search may become slower if evaluated at very long time series with
308 numerous neighbors. The reason is that all past sample data must be at first scanned at each
309 time step in order to locate the nearest neighbors, for which their corresponding errors are
310 then computed and the resulting error distribution is inferred (for each time step). Not until
311 then random samples from the derived error space can be picked up. Therefore, as we tested
312 our approach on very long time series while in real world applications less data are usually
313 available, the feasibility and usefulness of the method are confirmed. Nevertheless,
314 depending on modeling purposes, a compromise should be sought between the opposite
315 needs to consistently describe the model error and reduce the computational burden.

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

320 The above proposed technique for sampling from the model error e is based on the
321 assumption that a consistent description of the statistical properties of e can be provided by a
322 sufficiently long sample of model errors themselves that were experienced in validation.
323 Noteworthy, similar assumptions have been recently questioned on the argument that
324 epistemic uncertainty, which affects hydrological models, cannot be represented statistically
325 in view of the fact that disinformative data and epistemic error can lead to short-term non-
326 stationarity in the error statistics that cannot be easily represented by a formal statistical error
327 model with constant parameters (Beven and Westerberg 2011; Beven and Smith 2013, this
328 issue). In our opinion, this line of thought, which implies that epistemic uncertainty is not

329 subject to probabilistic description, may be misleading. Within probability theory, the reason
330 that we use the concept of a random variable is that the quantity of interest is not
331 deterministically known. If a variable is affected by uncertainty, then it is modeled as a
332 random variable, irrespective of the origin of uncertainty. That is, it can be modeled by using
333 stochastics, even if the stochastic dynamics has been imposed due to epistemic uncertainty.
334 This latter may imply the presence of autocorrelation, heteroscedasticity or non-stationarity.
335 Actually, all these are nothing else than stochastic concepts whose definitions are formulated
336 within a stochastic framework. Therefore, invoking these properties to argue about
337 inappropriateness of a stochastic modeling framework is a logical inconsistency, in our view.
338 These properties may increase predictive uncertainty and may underline the need for longer
339 data series for performing statistical inference, but they do not prevent the application of
340 statistical (or data driven) approaches. Therefore, the presence of epistemic uncertainty may
341 affect the results of the proposed approach but does not affect its theoretical validity and does
342 not prevent its practical application.

343

344 **Case study and experiment setting**

345 **Study catchment and data**

346 The presented approach was tested on the catchment of the Secchia River (northern
347 Italy), which is a tributary to the Po River. The contributing area of the analyzed basin to the
348 control section which is conventionally located at Bacchello Bridge (62 km upstream of the
349 confluence in the Po River) is about 1214 km² and the length of the stream is of 98 km
350 (Montanari and Koutsoyiannis 2012). The altitude varies from about 30 m to 2121 m above
351 sea level. The mean annual precipitation for this catchment ranges between 700 and more
352 than 2000 mm per year (Montanari 2005).

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

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

371 **Generation of the synthetic data**

372 Synthetic precipitation data of 50 years on five rain gauges located within the
373 catchment were generated using the generalized multivariate Neyman-Scott rectangular
374 pulses model (Cowpertwait 1995) which was previously calibrated on the available recorded
375 data from two years (1972 and 1973; (Montanari 2005)). The hourly mean areal rainfall over

376 the basin was calculated as a sum of the rescaled simulated precipitation at five rain gauges
377 using the weights of their contributing polygons with the Thiessen method.

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

383 The generated 50-year synthetic data of precipitation and temperature were next used to
384 obtain synthetic river flow records as outputs from the rainfall-runoff model ADM (Franchini
385 1996) validated for this catchment. The lumped nine-parameters ADM model was calibrated
386 in previous study against historical data of the Secchia River (1972 and part of 1973) giving
387 satisfying goodness of fit; the Nash-Sutcliffe efficiency (Nash and Sutcliffe 1970) for the
388 validation period (the second part of 1973) was equal to 0.81. Further information on the
389 synthetic data generation for the Secchia River case study and on the ADM model and its
390 calibration can be found in Montanari (2005). These synthetic data for all examined variables
391 are hereinafter referred to as the “actual” data.

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

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

404 **Hydrological model HyMod**

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

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

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

420 **Simulation procedure and prediction limits generation**

421 The proposed approach was tested on the synthetic data derived as described above.
422 The available 50 years dataset was split into three periods in order to:
423 (1) calibrate the HyMod (calibration period, years: 1-30), (2) infer the error model of HyMod

424 structural deficits in validation mode (error identification period, years: 31-40) and (3) fully
425 validate the approach (testing period, years: 41-50).

426 The calibration of the HyMod was carried out by using a classical approach.
427 Specifically, we used the DREAM algorithm (Vrugt and Robinson 2007) to minimize the
428 sum of the squared residuals in the simulation of the synthetic data over the first 30 years
429 (years: 1-30). The Nash-Sutcliffe efficiency was 0.93 for the validation when using years 31-
430 50. In principle, more rigorous likelihood functions could be used for model calibration
431 instead of least squares (see, for instance, the solutions proposed by Schoups and Vrugt
432 (2010) and Pianosi and Raso (2012)). However, they would imply the introduction of
433 additional parameters which we preferred to avoid in order to carry out a consistent
434 comparison with the results presented by Montanari and Koutsoyiannis (2012).

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

443 By sampling randomly from the posterior parameters distribution $f_{\Theta}(\Theta)$ we obtained
444 6000 realizations of the parameter set, which we then used over the validation period to test
445 our approach (years: 41-50). Sampling a parameter vector instead of single parameters
446 independently allows retaining the mutual dependencies between all parameters.

447 The years 31-40 were used to infer the space of the model structure errors E_i . To fully
448 explore the error space, we conducted the resampling technique at every time step of the

449 predicted river flow (testing period) as described above in section “The Nearest Neighbor
450 approach”.

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

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

465 **Assessment of the hydrological predictions reliability**

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

470 For the practical use of the NN approach, we carried out a sensitivity analysis of the
471 model response for a few different values of the k parameter. Specifically, we performed
472 1000 MC simulations for six different values of k , namely $k = 5, 10, 20, 30, 50, 100$, each

473 time computing the 95% prediction limits and their observation coverage, as well as the
474 percentage of values lying within the prediction limits, which theoretically should be 95%.

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

487

488 **Results**

489 **Diagnosis of simulated errors**

490 The diagnostic plot of the residuals simulated with the NN technique is provided in Fig. 2 and
491 compared to the residuals of the calibrated hydrological model (HyMod). A plot of residuals
492 versus simulated values (right panel in Fig. 2) shows that points are randomly scattered
493 around the value of 0. However, a correlation between the magnitude of residuals and the
494 values of simulated flows is observed. This is in agreement with the assumption underlying
495 the proposed NN technique, which relates model errors to the simulated (observed) river
496 flows. The autocorrelation of residuals is presented on the left panel in Fig. 2. Although the

497 non-parametric NN technique does not explicitly account for the error autocorrelation, this is
498 preserved in simulated errors. The autocorrelation of simulated residuals (Fig. 2, left bottom)
499 is noticeable and in agreement with residuals of the calibrated model (Fig. 2, left top). This
500 proves that the NN technique effectively simulates hydrological model errors by relating
501 errors to the simulated flows and thus indirectly accounting for their correlation (present in
502 simulated flows).

503 **Model prediction efficiency**

504 Correcting simulated river flow via the NN method slightly improved the model
505 predictions (see Fig. 3). The corresponding Nash-Sutcliffe efficiency was found to be 0.83
506 against 0.82 (without modifying river flows) for the best prediction (mode), while RMSE was
507 reduced by about 27% (from 14.4 to 10.4 m³ s⁻¹). The improvements are especially visible for
508 the peaks, which are of the highest concern in flood predictions and preventions. Note,
509 however, that the main objective of the method is not the improvement of the prediction
510 accuracy, but the conversion of a deterministic prediction into probabilistic one by providing
511 confidence limits for the predicted variable.

512 **Sensitivity to the k value of the NN approach**

513 The sensitivity analysis proved that the k value, among the reasonable range between 5
514 and 100 considered, has very little influence on the simulated river flows and the
515 corresponding prediction limits. The prediction limits derived with different values of k
516 varied by less than 0.1% of the total observation coverage. This can be explained, first, by the
517 goodness of the model fit to the “actual” data and therefore similar model errors (deviations
518 between simulated and “actual” river flows) over the entire river flows. And second, in the
519 case of a training data set with a sufficient length (covering fully the river flow variation), the
520 differences in model errors estimated by assigned nearest neighbors for different k may be

521 considered statistically indifferent. The explanation is that when k is much smaller than the
522 calibration series length, it is always possible to find within the calibration set at least k well-
523 fitting neighbors. Therefore, to minimize the computation effort, we limited our analysis to
524 the $k = 10$ case. The sensitivity of the predictions to the k -value may, however, need to
525 increase in a situation of limited data for the NN search or when the model does not explain
526 satisfactorily the behavior of a catchment.

527 **Percentage coverage of the prediction limits**

528 The 95% prediction limits computed by MC simulation are shown in Fig. 4. These
529 cover 97.9% of all observations for the testing period against the theoretical value of 95%.
530 That means that the intervals between prediction limits are slightly overestimated. In
531 particular, 0.5% and 1.6% of the “actual” points lie above and below the prediction limits
532 respectively against the theoretical values of 2.5% for each of the intervals. This is a
533 satisfying achievement and a noticeable improvement in comparison to the recent study of
534 Montanari and Koutsoyiannis (2012) on the same catchment but with a different method of
535 modifying simulated river flow (the Meta-Gaussian approach by Montanari and Brath
536 (2004)), where 5.4% and 4.3% of the “actual” river flows fell, respectively, out of the upper
537 and lower 95% prediction limits.

538 **Verification of the coverage probability**

539 The coverage probability of the predictions was evaluated for 1000 repetitions of the
540 simulated river flows over 10 years of the validation period for $k = 10$. Figure 5 presents the
541 resulting coverage probability plot (CPP) for the case study (gray line). As can be seen from
542 the figure, the confrontation of the computed quantiles with the theoretical ones indicates
543 satisfactory predictions of the variable; the points lie along and close to the bisector line (1:1),
544 especially for lower quantiles, whereas, a slight overestimation of the predicted river flow for

545 higher quantiles is observed. Thus, the derived prediction limits may be considered as reliable
546 in flood forecasting. This is again a significant improvement in comparison with the previous
547 study of Montanari and Koutsoyiannis (2012), where the predictions were visibly more
548 underestimated for all quantiles (compare to the black line in Fig. 5).

549

550 **Concluding remarks**

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

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

567 The results confirm that error resampling techniques may be an interesting option to
568 account for prediction uncertainty thereby avoiding a formal statistical characterization of the
569 model error, when it is difficult to parameterize. The proposed approach presents the

570 limitation of requiring a sufficiently long enough data records, first, for the hydrological
571 model calibration and second, for the error characterization using the nearest neighbor
572 technique. In order to provide reliable estimations, the same data set cannot be used twice.

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

580 **Acknowledgements**

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

590

591 **References**

592 Ajami, N. K., Duan, Q., and Sorooshian, S. (2007). “An integrated hydrologic Bayesian
593 multimodel combination framework: Confronting input, parameter, and model

594 structural uncertainty in hydrologic prediction.” *Water Resources Research*,
595 43(W01403), 19 pp.

596 Beven, K.J. (2006). "A manifesto for the equifinality thesis", *Journal of Hydrology*, 320, 18-
597 36.

598 Beven, K.J. (2010). "Preferential flows and travel time distributions: defining adequate
599 hypothesis tests for hydrological process models", *Hydrological Processes*, 24, 1537–
600 1547).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

666 Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M. (2012). "Towards a
667 comprehensive assessment of model structural adequacy." *Water Resources Research*,
668 48(W08301), 16 pp.

669 Hajebi, K., Abbasi-Yadkori, Y., Shahbazi, H., and Zhang, H. (2011). "Fast approximate
670 nearest-neighbor search with k-nearest neighbor graph." *Proceedings of the Twenty-
671 Second international joint conference on Artificial Intelligence-Volume Volume Two*,
672 1312–1317.

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

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

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

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

682 Koutsoyiannis, D., (2013). "Hydrology and Change", *Hydrological Sciences Journal*, 58(6),
683 1177-1197, doi:10.1080/02626667.2013.804626.

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

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

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

692 Laio, F., and Tamea, S. (2007). "Verification tools for probabilistic forecasts of continuous
693 hydrological variables." *Hydrology and Earth System Sciences Discussions*, 11(4),
694 1267–1277.

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

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

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

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

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

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

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

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

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

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

721 Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L.,
722 Koutsoyiannis, D., Cudennec, C., Toth, Grimaldi, S., Blöschl, G., Sivapalan, M.,
723 Beven, K., Gupta, H., Hipsey, M., Schaeffli, B., Arheimer, B., Boegh, E., Schymanski,
724 S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D.,
725 Srinivasan, V., Harman, C., Thompson, S., Rogger, Viglione, A., McMillan, H.,
726 Characklis, G., Pang, Z., and Belyaev, V. (2013). "'Panta Rhei—Everything Flows":
727 Change in hydrology and society—The IAHS Scientific Decade 2013–2022",
728 *Hydrological Sciences Journal*, 58 (6), doi: 10.1080/02626667.2013.809088.

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

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

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

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

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

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

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

747 Neuman, S. P. (2003). "Maximum likelihood Bayesian averaging of uncertain model
748 predictions." *Stochastic Environmental Research and Risk Assessment*, 17(5), 291–
749 305.

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

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

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

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

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

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

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

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

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

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

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

785 *Symposium, Newcastle: Adapting to Change.*, Barton, A.C.T.: Engineers Australia,
786 2009, 1028–1039.

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

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

793 Toth, E., Montanari, A., and Brath, A. (1999). "Real-time flood forecasting via combined use
794 of conceptual and stochastic models." *Physics and Chemistry of the Earth (B)*, 24(7),
795 793–798.

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

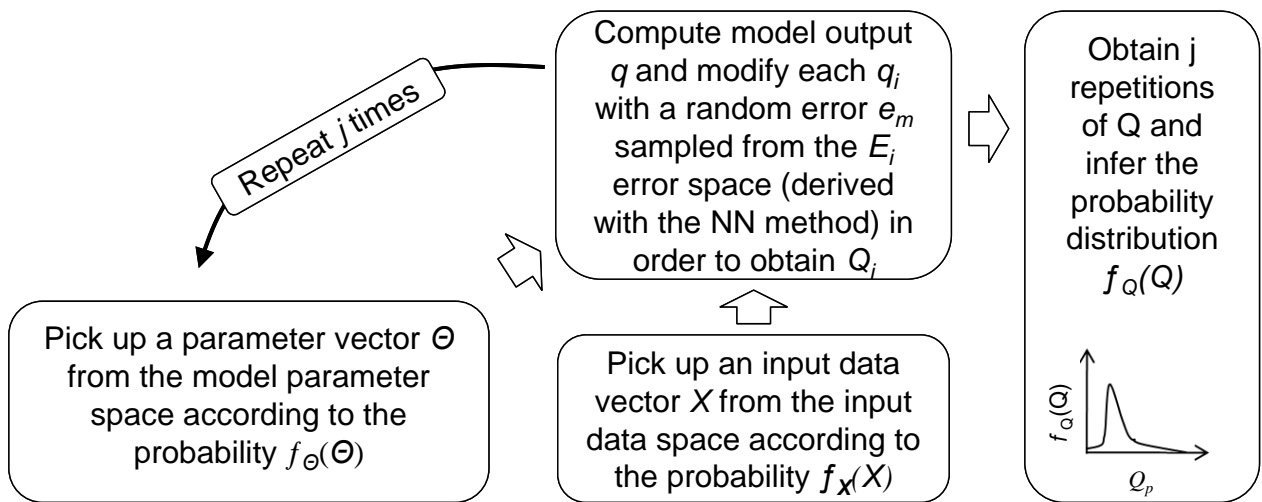
799 Yang, J., Reichert, P., Abbaspour, K. C., and Yang, H. (2007). "Hydrological modelling of
800 the Chaohe Basin in China: Statistical model formulation and Bayesian inference."
801 *Journal of Hydrology*, 340(3-4), 167–182.

802 Winsemius, H.C., Savenije, H.H.G., Gerrits, A.M.J., Zapreeva, E.A., and Klees, R. (2006).
803 "Comparison of two model approaches in the Zambezi river basin with regard to
804 model reliability and identifiability." *Hydrology and Earth System Sciences*, 10: 339--
805 352.

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

809 Winsemius, H.C., Schaefli, B., Montanari, A., Savenije, H.H.G. (2009). "On the calibration
810 of hydrological models in ungauged basins: A framework for integrating hard and soft
811 hydrological information". *Water Resources Research*, 45: doi:
812 10.1029/2009WR007706,2009.

813



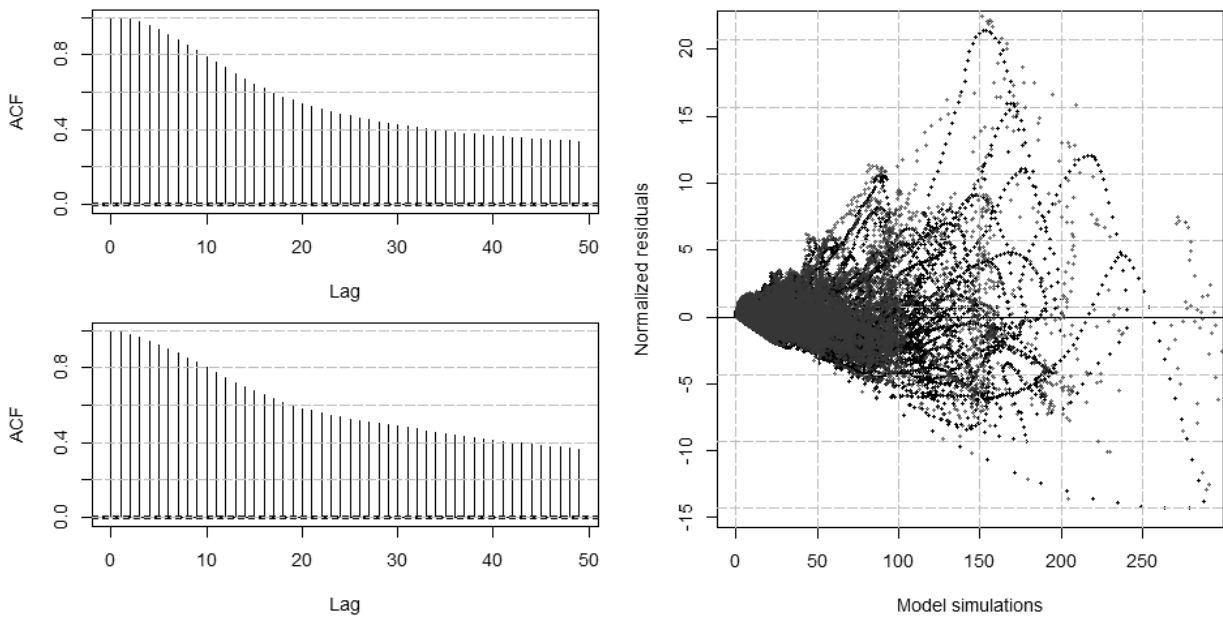
814

815

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

817

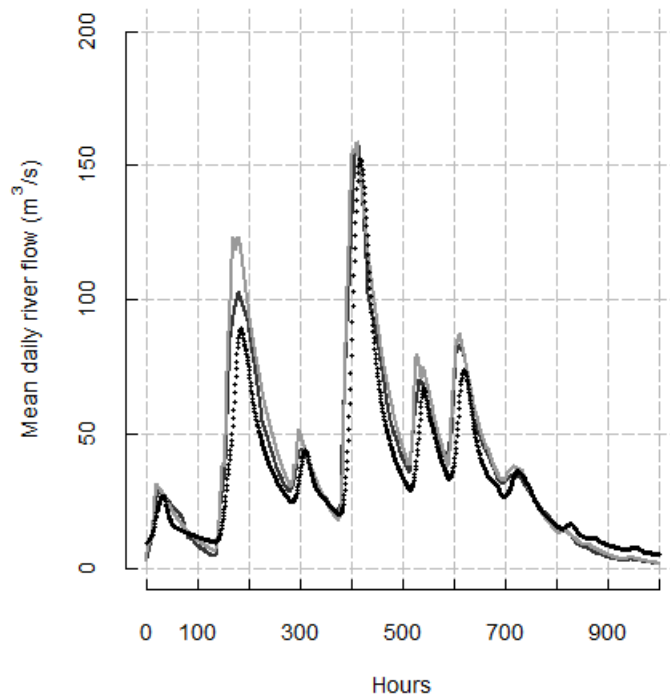
818



819
820

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

825

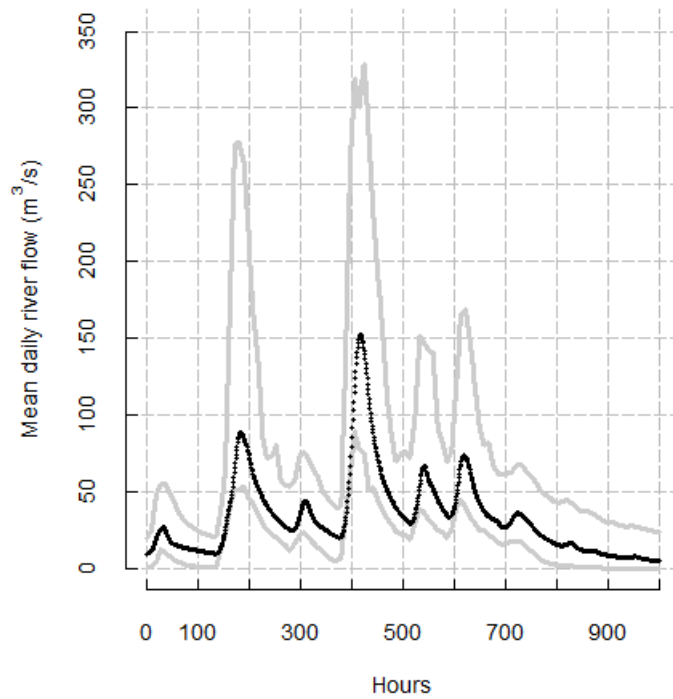


827

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

832

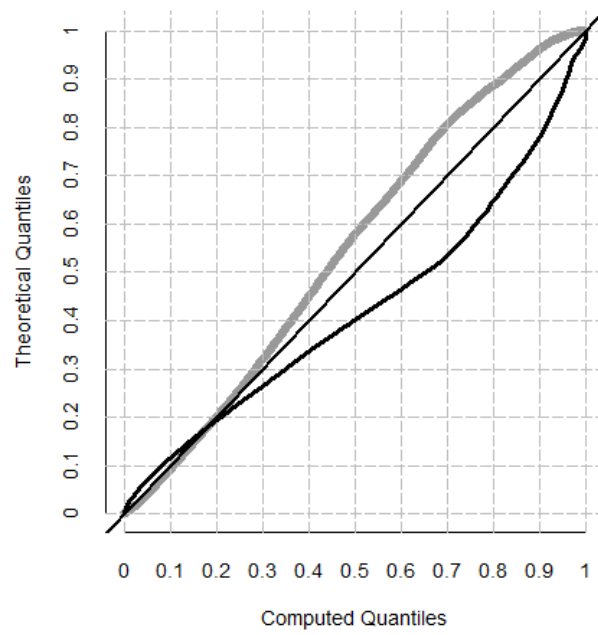
833



834

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

838



840

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