A Blueprint for Physically-Based Modeling of Uncertain Hydrological Systems

Alberto Montanari[1] and Demetris Koutsoyiannis[2]

A. Montanari, Department DICAM, University of Bologna, via del Risorgimento 2, Bologna, I-40136, Italy. (alberto.montanari@unibo.it)

D. Koutsoyiannis, Department of Water Resources and Environmental Engineering, National Technical University of Athens, Heroon Polytechneiou 5, Zographou, GR-157 80, Greece.

[1]Department DICAM, University of Bologna, Bologna, Italy.

[2]Department of Water Resources and Environmental Engineering National Technical University of Athens, Zographou, Greece.

Abstract. We present a new methodological scheme for building physically-based models of uncertain hydrological systems, thereby unifying hydrological modeling and uncertainty assessment. This scheme accounts for uncertainty by shifting from one to many applications of the selected hydrological model, thus formalizing what is done by several procedures for uncertainty estimation. We introduce a probability based theory to support the new blueprint and to ensure that uncertainty is efficiently and objectively represented. We discuss the related assumptions in detail, as well as the open research questions. We also show that the new blueprint includes as special cases the uncertainty assessment methods that are more frequently used in hydrology. The theoretical framework is illustrated by presenting a real-world application. In our opinion, the new blueprint could contribute to setting up the basis for a unified theory of uncertainty assessment in hydrology.

1. Introduction

Physically-based modeling has been a major focus for hydrologists for four decades already. In fact, more than forty years passed since *Freeze and Harlan* [1969] proposed their "physically-based digitally simulated hydrologic response model". An excellent review of the related research activity during the following thirty years was presented by *Beven* [2002]. Perhaps the most known physically-based model in hydrology is the Système Hydrologique Européen (SHE, see *Abbot et al.* [1986]), which has been the subject of many contributions. In the past ten years, physically-based modeling has been one of the targets of the well known "Prediction in Ungauged Basins" (PUB; see *Kundzewicz* [2007]) initiative of the International Association of Hydrological Sciences (IAHS). However, during the last four decades it became increasingly clear that uncertainty inherent to hydrological processes may make the use of a deterministic model inappropriate (see, for instance, *Grayson et al.* [1992]; *Beven* [1989, 2001]).

In fact, parallel research activity has shown the prominent role of uncertainty in hydrological modeling. Some authors expressed their belief that uncertainty in hydrology is epistemic and therefore can be in principle eliminated through a more accurate physical representation of the related processes [*Sivapalan et al.*, 2003]. However, recent contributions suggested that uncertainty is unavoidable in hydrology, originating from natural variability and inherent randomness (see, for instance, *Montanari et al.* [2009]; *Koutsoyiannis* [2010]). As a matter of fact, the presence of uncertainty makes the use of deterministic models impossible.

Given that traditionally physically-based models are built through deterministic equations, the above emerging limitations of deterministic models may induce one to conclude that fully physically-based models are not a feasible target in hydrology, because of their incapability to deal with uncertainty [*Beven*, 2002]). Therefore two relevant questions can be raised: is it possible to cope with uncertainty while retaining a physically-based approach? And, is it possible to perform physically-based hydrological modeling and uncertainty assessment within a unified theoretical framework?

We argue that the reply to both questions above is "Yes". In agreement with *Beven* [2002], we believe that a new blueprint should be established to overcome the incapability of traditional physically-based models to cope with uncertainty. We propose that the new blueprint is built on a key concept that is actually well known: it is stochastic physically-based modeling, which needs to be brought to a new light in hydrology. Here the term "stochastic" is used to collectively represent probability, statistics and stochastic processes.

In the next Section of the paper we take some notes on terminology which also provide the rationale for stochastic physically-based modeling. The third section of the paper is dedicated to the theory underlying the new blueprint that we are proposing. The fourth section describes the practical application of the proposed blueprint. The fifth section reviews the underlying assumptions and their limitations. Open research questions are discussed in the sixth section while the seventh is dedicated to placing existing approaches to uncertainty assessment in hydrology within the new blueprint. Finally we present an example of application and draw some conclusions.

2. Some notes on Terminology

First, let us note that traditionally, the term "physically-based model" is at the same time indicating a "spatially-distributed model" and a "deterministic model". We believe that this is not correct and therefore provide the following clarifications for these terms.

A physically-based model builds on the application of the laws of physics. In view of the extreme complexity, diversity and heterogeneity of meteorologic and hydrological processes (rainfall, soil properties...) physically-based equations are typically (but not necessarily) applied at local (small spatial) scale, therefore implementing a spatially-distributed representation. Spatial discretization is obtained by subdividing the catchment in subunits (subcatchments, regular grids, or other discretization methods). On the other hand, we may note that a full "reductionist" approach, in which all heterogeneous details of a catchment would be modelled explicitly and the modeling of details would provide the behaviour of the entire system, is a hopeless task [*Savenije*, 2009]. Indeed, some degree of approximation is unavoidable [*Beven*, 1989].

In hydrology, the most used physical laws are the gravitation law of Newton and the laws of conservation of mass, energy and momentum. However, it may be useful to make two clarifications, here:

1. While these laws give simple and meaningful descriptions of problems in simple systems, their application in hydrological systems demands simplification, lumping and statistical parameterization, and sometimes even replacing by conceptual or statistical laws (e.g. the Manning formula. See *Beven* [1989] for an extended discussion).

2. Hydrometeorological processes are governed by the laws of thermodynamics, which are, au fond, statistical physical laws. In this respect, complex systems cannot be modelled without enrolling statistics, as an inextricable part of physics.

These above arguments are usually forgotten and thus physically-based models typically refer to models reducible to Newton's and conservation laws. However, in this case one could conclude that a physically-based model is a delusion: even the simplest hydrological system is not reducible to such simple elements that these laws could be applied in their original form. For these reasons, here we use the term "physically-based model" with a wider content, so as to include some conceptualizations and statistical parameterizations. Furthermore, one may note that a hydrological model should, in addition to be physically-based, also consider chemistry, ecology, and so on (see, for instance, *Laio* [2006], *Hopp et al.* [2009]). We will focus here on physically-based models only, but the framework that we propose is generally applicable with any type of approach.

Many models in hydrology, including the physically-based ones, are often presented in deterministic form. Actually, a deterministic model is one where outcomes are precisely determined through known relationships among states and events, without any room for random variation. In such model, a given input will always produce the same output and therefore uncertainty is not taken into account. This is a relevant limitation, given that uncertainty is always present in hydrological processes, which is not just related to limited knowledge (epistemic uncertainty). It is rather induced, at least in part, by the above mentioned inherent variability and therefore it is unavoidable (see *Koutsoyiannis et al.* [2009]). It follows that deterministic representation, strictly speaking, is not possible in catchment hydrology. However, this conclusion does not apply to physically-based models which, in view of our reasoning above, are not necessarily deterministic.

There are many possible alternatives to deal with uncertainty thereby overcoming the limitations of deterministic approaches, including subjective approaches like fuzzy logic, possibility theory, and others [*Montanari*, 2007]. We believe that one of the most comprehensive ways of dealing with uncertainty is provided by the theory of probability. In fact, probabilistic descriptions allow predictability (supported by deterministic laws) and unpredictability (given by randomness) to coexist in a unified theoretical framework, therefore giving one the means to efficiently exploit and improve the available physical understanding of uncertain systems [*Koutsoyiannis et al.*, 2009]. The theory of stochastic processes also allows the incorporation into our descriptions of (possibly man induced) changes affecting hydrological processes [*Koutsoyiannis*, 2011], by modifying their physical representation and/or their statistical properties (see, for instance, *Merz and Blöschl* [2008a, b]). Finally, subjectivity and expert knowledge can be taken into account in prior distribution functions through Bayesian theory [*Box and Tiao*, 1973].

Therefore, a stochastic representation is a valuable opportunity in catchment hydrology, implying that a possible solution to model uncertain systems with a physically-based approach is the above mentioned stochastic physically-based modeling. We formalize the theoretical framework for the application of this type of approach here below.

3. Formulating a Physically-Based Model Within a Stochastic Framework

In this section we show how a deterministic model can be converted into an essential part of a wider stochastic approach through an analytical transformation, by simply introducing a deviation (an error term) from a single-valued relationship. The above mentioned analytical transformation is rather technical and is expressed by equations (1) to (6) below. We would like to introduce the new blueprint with a fully comprehensible treatment for those who are not acquainted with (or do not like) statistics. Therefore, the presentation is structured to allow the reader who is interested in the application only to directly jump to equations (7) and (8) without any loss of practical meaning.

Hydrological models are often expressed through a deterministic formulation, namely, a single valued transformation. In general, it can be written as

$$Q_p = S\left(\boldsymbol{\epsilon}, \mathbf{I}\right) \tag{1}$$

where Q_p is the model prediction which, in a deterministic framework, is implicitly assumed to equal the true value of the variable to be predicted. The mathematical relationship S represents the model structure, $\mathbf{I}$ is the input data vector and $\boldsymbol{\epsilon}$ the parameter vector. In the stochastic framework, the hydrological model is expressed in stochastic terms, namely [*Koutsoyiannis*, 2010],

$$f_{Q_p}\left(Q_p\right) = K f_{\boldsymbol{\epsilon},\mathbf{I}}\left(\boldsymbol{\epsilon}, \mathbf{I}\right) \tag{2}$$

where f indicates a probability density function, and K is a transfer operator that depends on deterministic model S. Within this context, Q_p indicates the true variable to be predicted, which is unknown at the prediction time and therefore is treated as a random variable.

Given that a single-valued transformation $S\left(\boldsymbol{\epsilon}, \mathbf{I}\right)$ as in eq. (1) represents the deterministic part of the hydrological model, the operator K will be similar to the Frobenius-Perron operator (e.g. *Koutsoyiannis* [2010]). However, K can be generalized to represent a so-called stochastic operator, which implements a shift from one to many transformations S.

A stochastic operator can be defined by using a stochastic kernel $k\left(e, \boldsymbol{\epsilon}, \mathbf{I}\right)$, with e reflecting a deviation from a single valued transformation. Let e be a stochastic process, with marginal probability density $f_e(e)$, representing the global model error according to the additive relationship

$$Q_p = S\left(\boldsymbol{\epsilon}, \mathbf{I}\right) + e \ . \tag{3}$$

Note that alternative error structures can be defined, for instance by introducing multiplicative terms. Here the global model error e is defined as the difference between the true value and the simulation provided by a given model with fixed parameters and input data.

The stochastic kernel introduced above must satisfy the following conditions:

$$k\left(e, \boldsymbol{\epsilon}, \mathbf{I}\right) \geq 0 \text{ and} \int_e k\left(e, \boldsymbol{\epsilon}, \mathbf{I}\right) de = 1 \ , \tag{4}$$

which are met if $k\left(e, \boldsymbol{\epsilon}, \mathbf{I}\right)$ is a probability density function with respect to e.

Specifically, the operator K applying on $f_{\boldsymbol{\epsilon},\mathbf{I}}\left(\boldsymbol{\epsilon}, \mathbf{I}\right)$ is then defined as [*Lasota and Mackey*, 1985, p. 101]

$$K f_{\boldsymbol{\epsilon},\mathbf{I}}\left(\boldsymbol{\epsilon}, \mathbf{I}\right) = \int_{\boldsymbol{\epsilon}} \int_{\mathbf{I}} k\left(e, \boldsymbol{\epsilon}, \mathbf{I}\right) f_{\boldsymbol{\epsilon},\mathbf{I}}\left(\boldsymbol{\epsilon}, \mathbf{I}\right) d\boldsymbol{\epsilon} d\mathbf{I} \ . \tag{5}$$

If the vectors of random variables $\boldsymbol{\epsilon}$ and $\mathbf{I}$ are independent to each other (although dependence may be present among the components of each one), the joint probability distribution $f_{\boldsymbol{\epsilon},\mathbf{I}}\left(\boldsymbol{\epsilon}, \mathbf{I}\right)$ can be substituted by the product of the two marginal distributions $f_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}\right) f_{\boldsymbol{I}}\left(\boldsymbol{I}\right)$. In view of this latter result, by combining eq. (2) and eq. (5), in which the model error can be written as $e = Q_p - S(\boldsymbol{\epsilon}, \mathbf{I})$ according to eq. (3), one obtains

$$f_{Q_p}(Q_p) = \int_{\boldsymbol{\epsilon}} \int_{\mathbf{I}} k\left[Q_p - S(\boldsymbol{\epsilon}, \mathbf{I}), \boldsymbol{\epsilon}, \mathbf{I}\right] f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) f_{\mathbf{I}}(\mathbf{I})\, d\boldsymbol{\epsilon} d\mathbf{I}\,. \tag{6}$$

At this stage one needs to identify a suitable expression for $k\left[Q_p - S(\boldsymbol{\epsilon}, \mathbf{I}), \boldsymbol{\epsilon}, \mathbf{I}\right]$. Upon substituting eq. (3) in eq. (6) and remembering that k is a probability density function with respect to the global model error e, we recognize that the kernel is none other than the conditional density function of e for the given $\boldsymbol{\epsilon}$ and $\mathbf{I}$, i.e., $f_{e|\boldsymbol{\epsilon},\mathbf{I}}\left[Q_p - S(\boldsymbol{\epsilon}, \mathbf{I}) | \boldsymbol{\epsilon}, \mathbf{I}\right]$.

To summarise the whole set of analytical derivations expressed by equations (1) to (6) one may conclude that we passed from the deterministic formulation of the hydrological model expressed by eq. (1), which we replicate for clarity here below,

$$Q_p = S(\boldsymbol{\epsilon}, \mathbf{I}) \tag{7}$$

to the stochastic formulation expressed by

$$f_{Q_p}(Q_p) = \int_{\boldsymbol{\epsilon}} \int_{\mathbf{I}} f_{e|\boldsymbol{\epsilon},\mathbf{I}}\left[Q_p - S(\boldsymbol{\epsilon}, \mathbf{I}) | \boldsymbol{\epsilon}, \mathbf{I}\right] f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) f_{\mathbf{I}}(\mathbf{I})\, d\boldsymbol{\epsilon} d\mathbf{I} \tag{8}$$

with the following meaning of the symbols:

- $f_{Q_p}(Q_p)$: probability density function of the true value of the hydrological variable to be predicted;

- $S(\boldsymbol{\epsilon}, \mathbf{I})$: deterministic part of the hydrological model;

- $f_{e|\boldsymbol{\epsilon},\mathbf{I}}\left[Q_p - S(\boldsymbol{\epsilon}, \mathbf{I}) | \boldsymbol{\epsilon}, \mathbf{I}\right]$: conditional probability density function of the global model error. According to eq. 2 it can also be written as $f_{e|\boldsymbol{\epsilon},\mathbf{I}}(e|\boldsymbol{\epsilon}, \mathbf{I})$;

- $\boldsymbol{\epsilon}$: model parameter vector;

- $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})$: probability density function of model parameter vector;

- $\mathbf{I}$: input data vector;

- $f_{\mathbf{I}}(\mathbf{I})$: probability density function of input data vector.

In eq. (8) the conditional probability distribution of the global model error $f_{e|\boldsymbol{\epsilon},\mathbf{I}}\left[Q_p - S\left(\boldsymbol{\epsilon},\mathbf{I}\right)|\boldsymbol{\epsilon},\mathbf{I}\right]$ is conditioned on the input data vector $\mathbf{I}$ and the parameter vector $\boldsymbol{\epsilon}$. Such formulation would be useful if one needed to account for changes in time of the conditional statistics of the model error (like, for instance, those originated by heteroscedasticity). On the other hand, if one assumed that the global model error is independent of $\mathbf{I}$ and $\boldsymbol{\epsilon}$, then eq. (8) can be written in the simplified form

$$f_{Q_p}\left(Q_p\right) = \int_{\boldsymbol{\epsilon}}\int_{\mathbf{I}} f_e\left[Q_p - S\left(\boldsymbol{\epsilon},\mathbf{I}\right)\right] f_{\boldsymbol{\epsilon}}\left(\boldsymbol{\epsilon}\right) f_{\mathbf{I}}\left(\mathbf{I}\right) d\boldsymbol{\epsilon} d\mathbf{I}\,. \tag{9}$$

We anticipate that one of the key issues is to efficiently represent the statistical properties of the global model error, as many contributions proposed by the hydrological literature already pointed out (see, for instance, *Refsgaard et al.* [2006]; *Kuczera et al.* [2006]; *Beven* [2006]).

The presence of a double integral in eq. (8) and eq. (9) may induce the feeling in the reader that the practical application of the proposed framework is cumbersome. Actually, the double integral can be easily computed through numerical integration, namely, by applying a Monte Carlo simulation procedure that is well known and already used in hydrology (see *Koutsoyiannis* [2010]). We explain the numerical integration in the next section of the paper.

4. Application of the Proposed Framework: Joining Hydrological Model Implementation and Uncertainty Assessment

Estimating the probability distribution of the true value of the variable to predicted by a hydrological model is equivalent to simultaneously carry out model implementation and uncertainty assessment. The framework for estimating the probability density function of model prediction, $f_{Q_p}(Q_p)$, was proposed in Section 3. Here we show how eq. (8) can be applied in practice.

Let us admit that the hydrological model is fully generic and possibly physically-based. Also, let us admit that the probability density functions of model parameters, input data and model error are known, for instance because they were already estimated by using procedures that were proposed by the hydrological literature (see, for instance, *Di Baldassarre and Montanari* [2009] for input data uncertainty, *Vrugt et al.* [2007] for parameter uncertainty and *Montanari and Brath* [2004] for global model error). A practical demonstration showing how this can be determined is contained in Section 8 below.

Under the above circumstances the double integral in eq. (8) can be easily computed through a Monte Carlo simulation procedure, which can be carried out in practice by performing many implementations of the deterministic hydrological model $S(\boldsymbol{\epsilon}, \mathbf{I})$.

In detail the simulation procedure is carried out through the following steps:

1. A parameter vector for the hydrological model is picked up at random from the model parameter space according to the probability distribution $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})$.

2. An input data vector for the hydrological model is picked up at random from the input data space according to the probability distribution $f_{\mathbf{I}}(\mathbf{I})$.

3. The hydrological model is run and a model prediction (or a vector of individual predictions) $S(\boldsymbol{\epsilon}, \mathbf{I})$ is computed.

4. A number n of realizations of the global model error (or vectors of individual errors) is picked up at random from the model error space according to the probability distribution $f_{e|\boldsymbol{\epsilon},\mathbf{I}}(e)$ and added to the model prediction $S(\boldsymbol{\epsilon}, \mathbf{I})$.

5. The simulation described by items from 1 to 4 is repeated j times. Therefore one obtains $n \cdot j$ (vectors of) realizations of the true variable to be predicted Q_p.

6. Finally the probability distribution $f_{Q_p}(Q_p)$ is inferred through the realizations mentioned in item 5.

It is important to note that j needs to be sufficiently large, in order to accurately estimate the probability density $f_{Q_p}(Q_p)$. It is also clarified that in a typical Monte Carlo procedure one would use n = 1 (where, to each simulation a different realisation of model error e would be generated). However, a larger n value multiplies the number of simulated points by a factor n with negligible increase of computer time (as the same hydrological simulation run is used for all n). We believe that a modest value of n (see application in Section 8) results in a good compromise of accuracy and computational efficiency. Figure 1 shows a flowchart of the whole simulation procedure.

Once the probability distribution of the true value to be predicted Q_p is known the problems of hydrological modeling and uncertainty assessment are both solved.

5. Discussion of the Underlying Assumptions

Like any scientific method, the blueprint proposed in Section 3 and 4 is based on assumptions in order to ensure applicability. When dealing with uncertainty assessment in hydrology, assumptions are often treated with suspect, because it is felt that they undermine the effectiveness of the method and therefore its efficiency and credibility with respect to users. We must admit, though, that assumptions are unavoidably needed to set up models, calibrate their parameters and estimate their reliability, whatever approach is used. Evidently, flawed assumptions may falsify statistical inference as well as any alternative model of uncertain and deterministic systems. Therefore the target of the researcher should not be to avoid assumptions, but rather discuss them transparently, evaluate their effects and, when possible, check them, for instance through statistical testing.

In order to discuss the assumptions conditioning the blueprint we introduced above, first note that the theoretical scheme is very general. In fact, we only assumed that model input data and parameters are random vectors which are independent to each other. Such assumption implies that parameter uncertainty is independent of data uncertainty and its influence on the results depends on data uncertainty itself. Unless this latter is very significant, we believe the assumption is reasonable. In principle the above assumption could be removed by estimating the joint probability distribution of input data and parameters and then picking up from this distribution the random outcomes at steps 1 and 2 of the simulation procedure described in Section 4. Actually, statistical inference of joint probability distributions of model input data and parameters is likely to be affected by much uncertainty and therefore it may be more difficult to implement. We plan to study this solution in future research.

One may note that further assumptions might be needed to estimate the probability distribution f_e of model error, which might be non-Gaussian and affected by heteroscedasticity. For instance, in the application presented in Section 8 the meta-Gaussian approach by *Montanari and Brath* [2004] is applied. Actually, this method assumes that the joint probability distribution of model error and model simulation is stationary and independent of input uncertainty and parameter uncertainty, but the marginal probability distribution of the model error can eventually result heteroscedastic (see Section 8 and *Montanari and Brath* [2004]). If one used the Generalised Likelihood Uncertainty Estimation (GLUE; see *Beven and Binley* [1992]) different assumptions would be introduced depending on the (possibly informal) likelihood measure that is used to characterise the reliability of model output. No matter which method is used, any additional assumption introduced for inferring f_e should be appropriately checked.

A relevant issue has been pointed out by some authors (see, for instance, *Beven et al.* [2011]) who are convinced that epistemic errors arising from hydrological models might be affected by non-stationarity and therefore difficult (or impossible) to model by using statistical approaches. In our opinion epistemic uncertainty in itself, which is not changing in time, cannot induce non-stationarity, which might instead be necessary to enrol when environmental changes are present. However, independently from its origin, non-stationarity can be efficiently dealt with by using non-stationary stochastic processes, by introducing and checking suitable assumptions.

The conclusion of the above discussion can be summarised by saying that (a) the only relevant assumption conditioning the proposed blueprint is justified as long as data uncertainty is not very significant (see also additional discussion about this in Section 8.3). Within this respect, we would like to emphasise our opinion that in the presence of significant input data errors (also called "observation uncertainty") any uncertainty assessment method is ill-posed and likely to end up with underestimation. Moreover, (b) the above assumption can in principle be removed although it is likely that this option turns out to be more difficult to handle in practice. And finally, (c) further assumptions might be needed for ensuring the practical application of the approach, which should be appropriately checked.

6. Open Research Questions

The above discourse shows that to include a deterministic model within a stochastic framework is in principle possible. Although we explicitly focused on physically-based approaches, the blueprint that we are proposing is applicable to any deterministic scheme, therefore including conceptual and black-box models. We believe that incorporation of physically-based deterministic models bears a greater added value of the blueprint we are proposing. In fact, analyzing the randomness of physically-based systems is an invaluable opportunity to improve their understanding therefore increasing predictability, according to the "models of everywhere" concept [*Beven*, 2007].

However, relevant research challenges and practical problems may prevent a successful application of the blueprint. First of all, numerical integration (e.g. the Monte Carlo simulation outlined in Section 4) is computationally intensive and may result prohibitive for spatially-distributed models. Therefore efficient simulation schemes are necessary, while too detailed spatial representations may not make any difference except in wasting computer time.

Second, a relevant issue is the estimation of global model uncertainty, namely, the estimation of the probability distribution $f_e(e)$ of the model error. The literature has proposed a variety of different approaches, like the above mentioned GLUE method [*Beven and Binley*, 1992], the meta-Gaussian model [*Montanari and Brath*, 2004; *Montanari and Grossi*, 2008], Bayesian Model Averaging (BMA, *Neuman* [2003]) and BATEA [*Kuczera et al.*, 2006]. However, the above methods rely on limiting assumptions and some of them are too computer intensive. We believe that estimating global model uncertainty in hydrology [*Montanari*, 2011] is still an open problem for which more focused research is needed. The proposed framework may facilitate streamlining of this research and linking it with other components within an holistic modeling approach.

Finally, estimation of parameter uncertainty is a relevant challenge as well. Possibilities are the GLUE method [*Beven and Binley*, 1992] and the DREAM algorithm [*Vrugt et al.*, 2007], which nevertheless are computer intensive as well and may turn out to be inpractical with spatially-distributed models applied to fine time scale at large catchments.

7. Placing Uncertainty Assessment Techniques Within the Proposed Blueprint

The blueprint proposed in Section 3 and 4 aims to provide a general theoretical framework for uncertainty assessment in hydrology. Indeed, the most frequently used techniques can be easily placed within it. For instance, the well known GLUE method [*Beven and Binley*, 1992] anticipated many of the concepts we are highlighting here, and in particular the idea of estimating uncertainty by turning from one to many applications of the hydrological model. In detail, the simulation procedure used within the classical applications of GLUE is much similar to what is presented in Section 4. The only relevant difference is related to the estimation of global model error, which is resolved within GLUE by estimating the model likelihood, and therefore the probability distribution of the true variable to be predicted, through an integral performance measure or by fixing limits of acceptability [*Liu et al.*, 2009; *Winsemius et al.*, 2009]. In fact, likelihood is estimated within classical GLUE by adopting an informal approach, basing on a dummy likelihood measure (like the Nash-Sutcliffe efficiency in many GLUE applications). Basing on the blueprint proposed here, GLUE can then be defined as a statistical approach where model likelihood is estimated informally.

Moreover, the proposed blueprint reduces to the meta-Gaussian approach by *Montanari and Brath* [2004], once that parameter uncertainty and input uncertainty are neglected. A similar reasoning applies to the Bayesian Forecasting System by *Krzysztofowicz* [2002], where parameter uncertainty is neglected and the probability distribution of the true variable to be predicted is estimated by inferring the joint probability distribution of true value and corresponding model output.

8. An Example of Application

In order to illustrate the proposed blueprint with a practical example, an application is presented here below that refers to a rainfall-runoff model applied to a catchment located in Italy.

8.1. The study catchment

The application refers to the Leo River at Fanano, in the Emilia-Romagna region, in Northern Italy. Figure 2 shows the location of the catchment. The catchment area is 64.4 km^2 and the main stream length is about 10 km. The maximum elevation in the catchment is the Mount Cimone (2165 m a.s.l.), which is the highest peak in the northern part of the Apennine Mountains. The climate over the region is continental.

Daily river flow data at Fanano are available for the period January 1st, 2003 - October 26th, 2008, for a total of 2126 observations. For the same period, daily mean areal rainfall and temperature data over the catchment are available, as estimated by the Italian National Hydrographic Service basing on observation collected in nearby gauging stations.

The observations collected from January 1st 2003 to December 31st 2007 were used for calibrating the rainfall-runoff model, while the period September 1st 2007 - October

26th 2008 was reserved for its validation. We estimated the probability distribution of the model error by referring to the first year of the validation period (2007), in order to obtain a reliable assessment of $f_{e|\boldsymbol{\epsilon},\mathbf{I}}(e|\boldsymbol{\epsilon},\mathbf{I})$ in a real world application. Note that the general formulation of eq. (8) is used, thereby accounting for heteroscedasticity in the model error itself. Finally, the period January 1st 2008 - October 26th 2008 was reserved for testing, in full validation mode, the proposed blueprint (rainfall-runoff modeling and uncertainty assessment).

8.2. The rainfall-runoff model

The rainfall-runoff model is AFFDEF [*Moretti and Montanari*, 2007], a spatially-distributed grid-based approach where hydrological processes are described with physically-based and conceptual equations. In order to limit the computational requirements, and in view of the limited catchment area, the Leo river basin was described by using only one grid cell, therefore applying a lumped representation. AFFDEF was calibrated by using DREAM [*Vrugt et al.*, 2007], that is, a modified SCEM-UA global optimisation algorithm [*Vrugt et al.*, 2003]. The DREAM method makes use of population evolution like a genetic algorithm together with a selection rule to assess whether a candidate parameter set is to be retained. The sample of retained sets after convergence can be used to infer the probability distribution of model parameters. Herein, a number of 1000 parameter sets were retained, which indirectly determine the density function $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})$ of the parameter vector in a non-parametric empirical manner, fully respecting the dependencies between different parameters.

AFFDEF explained about 57% and 50% of the river flow variance in calibration and validation, respectively. Figure 3 reports a comparison during the validation period (2007 and 2008) between observed and simulated hydrographs. This latter was obtained by using the best parameter set according to explained variance during the calibration period. One can see that a significant uncertainty affects the model performances, which is unlikely merely due to lumping the model at catchment scale. We are interested in checking whether the proposed blueprint provides a consistent assessment of such uncertainty.

Finally, the probability distribution of the model error was inferred by using the meta-Gaussian approach by *Montanari and Brath* [2004]. In brief, the method recognizes that the error is affected by heteroscedasticiticy by accounting for the dependence of its conditional probability distribution on model prediction. In this way change of the statistical properties during time is efficiently modeled. The estimation of the joint probability distribution of model simulation (provided by AFFDEF by using the best parameter set in terms of explained variance during the calibration period) and error is carried out by preliminarily transforming the data to the Gaussian probability distribution. In the Gaussian domain a bivariate Gaussian distribution is finally estimated. The goodness-of-fit provided by the meta-Gaussian approach was checked by using the statistical tests described in *Montanari and Brath* [2004], where more details on the procedure can be found.

8.3. The simulation procedure

We assumed to neglect input data uncertainty because no information was available to infer the probability distribution of the available observations. This is an important limitation in many practical applications. In particular, input uncertainty is usually dominant in real time flash-flood forecasting, where input rainfall to a rainfall-runoff model is usually predicted to increase the lead time of the river flow forecasting. If a probabilistic prediction for rainfall is available then input uncertainty can be efficiently taken into account in the blueprint proposed above. In alternative, input uncertainty can be estimated by using expert knowledge or Bayesian procedures like BATEA [*Kuczera et al.*, 2006]. Given that the present application refers to a past period and excludes future forecast inputs, and since data series have been tested, it is reasonable to assume data certainty, with the awareness that we may slightly underestimate prediction uncertainty in this case.

The simulation procedure was performed by running AFFDEF during the 300-day validation period January 1st 2008 - October 26, 2008, for each of the $j = 1000$ parameter sets retained by DREAM. Then, $n = 100$ random outcomes from the probability distribution of the model error were added to each observation of the 1000 simulated data series, therefore obtaining $1000 \cdot 100$ simulations of the data value referred to each of the above 300 days, which allowed us to estimate the related probability distribution. The above j and n values were selected by estimating the number of sampling points to efficiently infer the shape of the related probability distributions.

Figure 4 shows the 95% confidence band of the model simulation, along with the corresponding observations. It can be seen that the results are physically meaningful and, in our opinion, confirm the efficiency of the proposed blueprint. In fact, the confidence bands are quite large as one would expect by looking at the performances of the model, which underline the presence of significant uncertainty. A number of data points are located outside the confidence bands as one would expect by considering that the band itself is drawn at the 95% confidence level.

9. Conclusions

A new blueprint is presented for formulating physically-based models of uncertain hydrological systems, which in effect means all hydrological systems. The main advantages of the proposed methodological scheme are that (a) hydrological modeling and uncertainty assessment are jointly carried out and (b) a general theoretical framework is elaborated for uncertainty estimation in hydrology, which includes as special cases the existing and most frequently used methods.

Basically, the blueprint proposes to incorporate deterministic hydrological models within a stochastic framework. This solution is suggested by our convincement that probability is the most efficient and objective technique for uncertainty assessment. Shifting from the deterministic to the stochastic formulation requires passing from one to many applications of the hydrological model. What we suggest is not new in practical applications and constitutes also the rationale underlying some of the existing uncertainty assessment methods like GLUE [*Beven and Binley*, 1992]. However, a comprehensive theoretical framework is proposed, along with a detailed discussion of the underlying assumptions, therefore allowing one to structure in a objective setting the application of hydrological models in order for uncertainty to be taken into account and estimated.

An application to an Italian river basin is presented for illustrating the introduced blueprint. Although a simplifying assumption was introduced to neglect data uncertainty and a lumped rainfall-runoff model was used, the case study shows that the proposed approach is efficient and physically meaningful.

We believe the theoretical framework introduced here may open new perspectives regarding modeling of uncertain hydrological systems. In fact, statistical analysis of uncertainty and predictability offers valuable indications to improve our understanding of real systems and better understand their (possibly) changing or shifting behaviors and their reaction to (human induced) changes. Last but not least, we believe that the proposed procedure is very useful for educational purposes, putting the basis for developing a unified theoretical basis for uncertainty assessment in hydrology.

Relevant research questions are still open. The proposed procedure is based on running multiple simulations and therefore it is computationally intensive. For this reason, application to very detailed spatial representation implemented on complex systems may require significant computational resources. Finally, estimation of parameter uncertainty, global model uncertainty and data uncertainty may represent relevant problems for some real world applications, for which additional and focused research is needed.

Acknowledgments. To authors are grateful to Francesco Laio, Keith Beven and Elena Montosi for providing very useful comments and help. A.M. was partially supported by the Italian government through the grant "Uncertainty estimation for precipitation and river discharge data. Effects on water resources planning and flood risk management".

References

Abbot, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen (1986), An introduction to the European Hydrological System - Systme Hydrologique Europen, SHE. 2. Structure of a physically-based, distributed modeling system., *J. Hydrol.*, *102*, 87–77.

Beven, K. J. (1989), Changing ideas in hydrology The case of physically-based models, *J. Hydrol.*, *105*, 157–172.

Beven, K. J., and A. M. and Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Proc.*, *6*, 279–298.

Beven, K. J. (2001), How far can we go in distributed hydrological modeling, *Hydrol. and Earth System Sci.*, *5*, 1–12.

Beven, K. J. (2002), Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system, *Hydrol. Proc.*, *16*, 189–206.

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

Beven, K. J. (2007), Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process, *Hydrol. and Earth System Sci.*, *11*, 460–467.

Beven, K. J., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), *Hydrol. Earth Syst. Sci. Discuss.*, **8**, 5355–5386.

Box, G. E. P., and G. C. Tiao (1973), Bayesian Inference in Statistical Analysis, *Addis-onWesley*, Boston, Massachusetts.

Di Baldassarre, G, and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. and Earth Sys. Sci.*, *1*3, 913–921.

Freeze, R. A., and R. L. Harlan (2008), Blueprint for a physically-based, digitally- simulated hydrologic response model, *J. Hydrol.*, *9*, 237–258.

Grayson, R. B., I. D. Moore, and T. A. McMahon, (1992), Physically-based hydrologic modeling. 2. Is the concept realistic?, *Water Resour. Res.*, *28*, 2659–2666.

Hopp, L., C. Harman, S. L. E. Desilets, C. B. Graham, J. J. McDonnell, and P. A. Troch (2009), Hillslope hydrology under glass: confronting fundamental questions of soil-water-biota co-evolution at Biosphere 2, *Hydrol. Earth Syst. Sci.*, *13*, 2105-2118.

Kundzewicz, Z. W. (2007), Predictions in Ungauged Basins: PUB Kick-off, Proceedings of the PUB Kick-off meeting held in Brasilia, 2022 November 2002, *IAHS Publ.* 309, 38–44.

Koutsoyiannis, D., C. Makropoulos, A. Langousis, S. Baki, A. Efstratiadis, A. Christofides, G. Karavokiros, and N. Mamassis (2009), HESS Opinions: "Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability", *Hydrol. Earth Syst. Sci.*, *13*, 247–257.

Koutsoyiannis, D. (2010), HESS Opinions "A random walk on water", *Hydrol. Earth Syst. Sci.*, *14*, 585–601.

Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics and uncertainty, *Journal of the American Water Resources Association*, *47*, 481–495.

Krzysztofowicz, R. (2002), Bayesian system for probabilistic river stage forecasting, J. of Hydrol., *268*, *16–40.*

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*, 161–177.

Laio, F. (2006), A vertically extended stochastic model of soil moisture in the root zone, *Water Resour. Res.*, *42*, doi:10.1029/2005WR004502.

Lasota, D. A., and M. C. Mackey (1985), Probabilistic properties of deterministic systems, Cambridge University Press.

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

Merz, R., and G. Blöschl (2008a), Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, *44*, doi:10.1029/2007WR006744.

Merz, R., and G. Blöschl (2008a), Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, *44*, doi:10.1029/2007WR006745.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *40*, doi:10.1029/2003WR002540.

Montanari, A., What do we mean by uncertainty? The need for a consistent wording about uncertainty assessment in hydrology (2007), *Hydrol. Proc.*, *2*1, 841–845.

Montanari, A., and G. Grossi (2008), Estimating the uncertainty of hydrological forecasts: A statistical approach, *Water Resour. Res.*, *44*, doi:10.1029/2008WR006897.

Montanari, A. , C. A. Shoemaker, and N. van de Giesen (2009), Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resour. Res.*, *45*, doi:10.1029/2009WR008471.

Montanari, A. (2011), Uncertainty of Hydrological Predictions, in: Peter Wilderer (ed.) Treatise on Water Science, Oxford Academic Press, *2*, 459–478.

Moretti., G., and A. Montanari (2007), AFFDEF: A spatially distributed grid based rainfall-runoff model for continuous time simulations of river discharge, *Environ. Mod. & Soft.*, *22*, 823–836.

Neuman, S.P. (2003), Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models, *Stochastic Environmental Resources Risk Assessment*, *17*, 291–305.

Refsgaard, J. C., J. P. van Der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, *29*, 1586–1597.

Savenije, H. H. G. (2009), HESS Opinions “The art of hydrology”, *Hydrol. Earth Syst. Sci.*, *13*, 157–161.

Sivapalan, M., K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiondo, P.E. O’Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*, 857–880.

Vrugt J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*, doi:10.1029/2002WR001642.

Vrugt, J. A., and B. A. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. of the Natl. Acad. of Sci. of USA*, *1*04, 708–711.

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

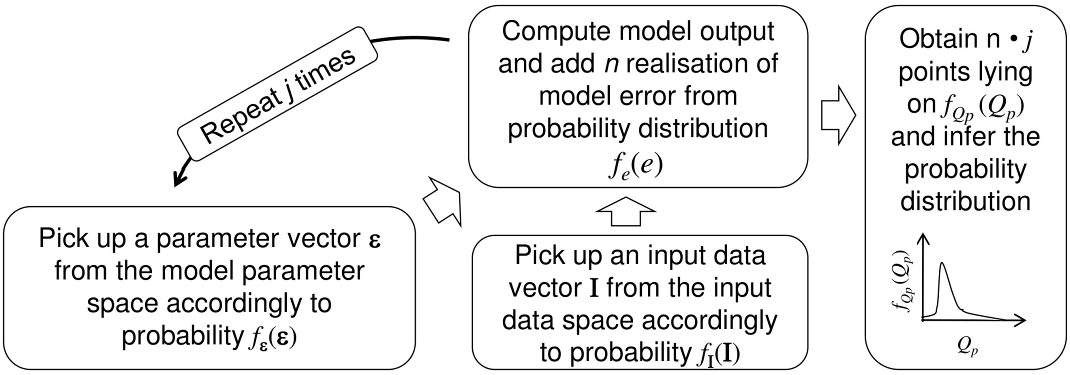

Figure 1. Flowchart of the Monte Carlo simulation procedure for performing the numerical integration in eq. (8). The marginal distribution f_e is replaced by the conditional $f_{e|\boldsymbol{\epsilon},\mathbf{I}}$ when the error depends on parameters and inputs.

Figure 2. Location of the Leo River basin (Italy).

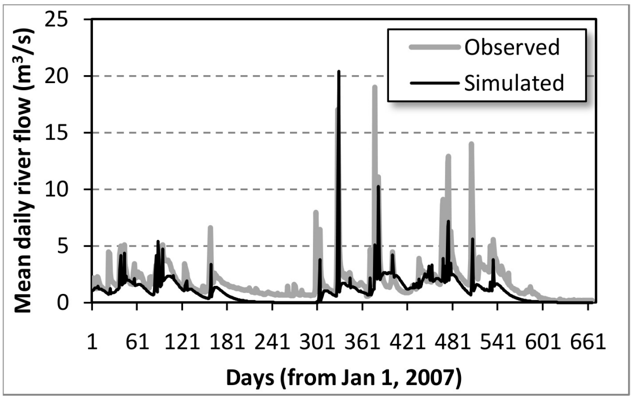

Figure 3. Comparison between observed and simulated hydrographs during the validation period (Jan 1st 2007 - October 26, 2008). The simulated hydrograph was obtained by using the best parameter set according to explained variance during the calibration period.

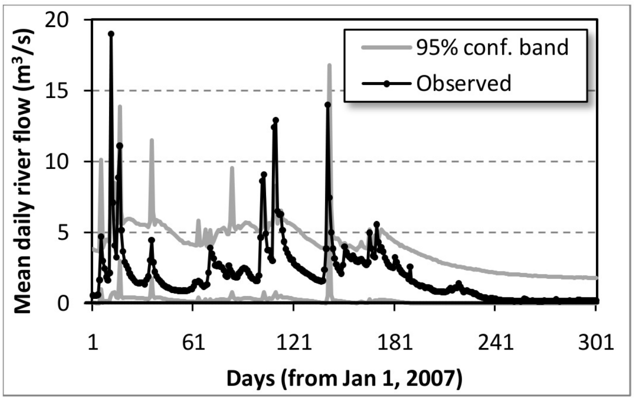

Figure 4. 95% confidence bands of the river flow simulation provided by AFFDEF during the validation period (Jan 1st, 2008 - Oct 26th, 2008) of the proposed blueprint, along with the corresponding observed values.