Prof. D.P. Solomatine, PhD, MSc LN0448/11/2

Uncertainty Analysis in Modelling

Uncertainty Analysis in Modelling

TABLE OF CONTENTS

This lecture notes present a brief coverage of the subject "Modelling theory and uncertainty" (mainly its second part, devoted to models uncertainty) given in the Masters course "Water Science and Engineering", Specialisation in Hydroinformatics.

1 Modelling process

1.1 Hydroinformatics: integration of data, models and humans

Hydroinformatics concerns the integrated use of

- information and communication technologies,
- modelling
- and decision support systems

in solving problems related to the aquatic environment and involving relevant stakeholders.

It studies the dynamics of information flow and the transformation of knowledge along the chain

"Data – models – knowledge – decisions – impacts".

More specifically, such transformation goes

- from data acquisition and analysis using advanced ICT and remote sensing tools;
- through computer-based modelling and forecasting of water motion and information flows with the appropriate uncertainty estimates;
- through multi-faceted optimisation of water management and control options, structures and models;
- all of which support managerial decisions and control actions, where hydroinformatics tools enhance social and stakeholder collaboration through participatory processes;
- and promote the uptake and dissemination of knowledge and the user's and stakeholders' feedback through appropriate use of communication systems like mobile telephony and Internet.

Hydroinformatics is based on a holistic view of water infrastructure and management issues, and creates the appropriate technological and institutional environments that contribute to social justice.

Hydroinformatics uses various forms of knowledge encapsulation:

- Tacit (implicit) knowledge embedded within a person
- Words, texts, images – printed, or stored in electronic media
- Mathematical models: formulas, algorithms encapsulated in computer programs (software).

Hydroinformatics systems are integrated systems encapsulating all of above (Figure 1).

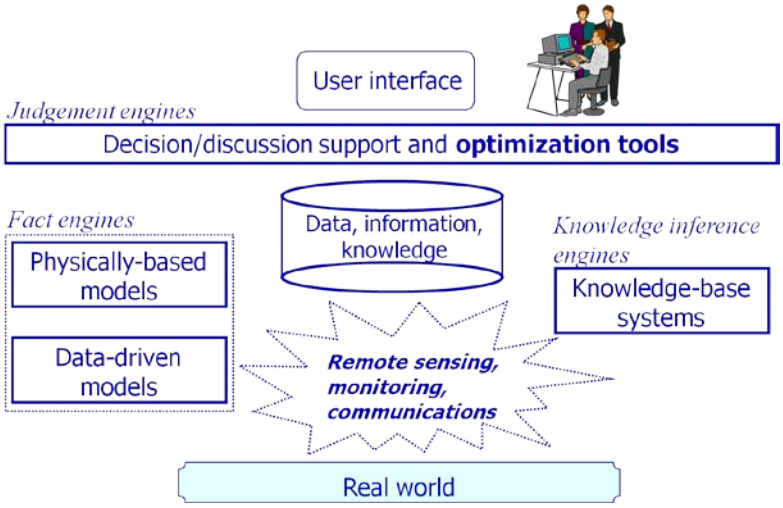


Figure 1. Hydroinformatics system.

Hydroinformatics deals with the flow of information over the hydrological cycle and the corresponding decision making and management processes – see Figure 2.

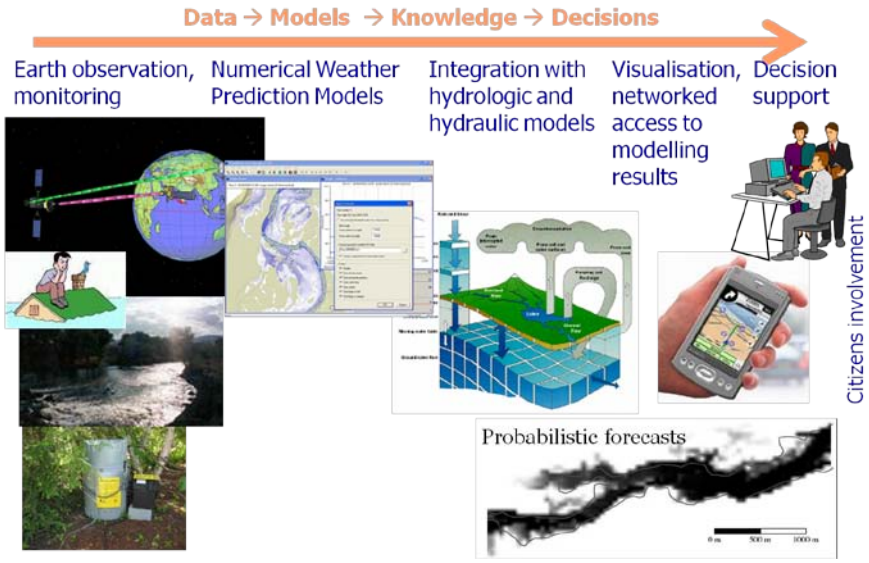


Figure 2. Flow of information related to hydrological cycle, to from data to decisions, handled by hydroinformatics

1.2 Modelling and the associated uncertainty

Modelling is an important part of Hydroinformatics, its main engine. Model is a simplified description of reality and encapsulates knowledge about a particular physical, social or mental process in electronic form. Goals of modelling are:

- understand the studied system or domain (understand the past)
- predict the future
- predict the future values of some of the system variables, based on the knowledge about other variables
- use results of modelling for making decisions (change future)

Two major modelling paradigms are distinguished:

- Physically-based model (also called process, simulation, knowledge-based, numerical) is based on the understanding of the underlying processes in the system
    - examples: river models based on main principles of water motion, expressed in differential equations, solved using finite-difference approximations
- Data-driven model is based on the recorded values of variables characterising the system. They need less knowledge about physical behaviour
    - examples: statistical model linking input and output

A typical mathematical model of a system (e.g. a river basin) is presented at Figure 1.

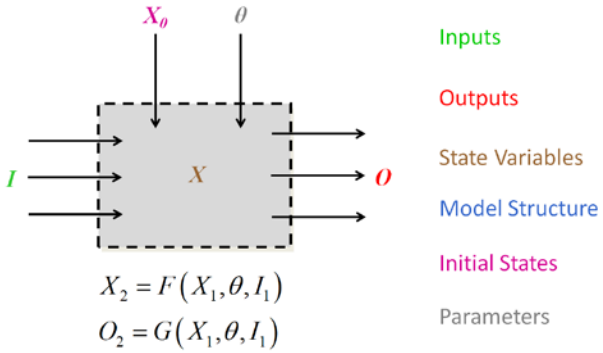


Figure 3. Dynamic mathematical model of a system.
I = inputs, O = outputs, X = state variables, X_0 = initial states, θ = parameters, t = time

Here

$X = F_1(I, \theta)$
$O = F_2(X, \theta)$ $O = F(I, \theta)$
in continuous time: $O(t) = F(I(t), \theta)$
in discrete time (step-wise change): $O(t+1) = F(I(t), \theta)$

Typical steps of modelling process are presented on Figure 2.

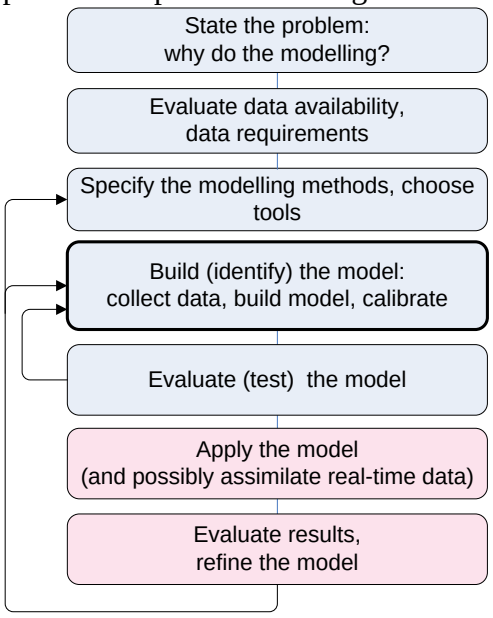


Figure 4. Steps in the modelling process

The following gives a bit more specifics to the diagram:

- State the problem (why do the modelling?)
- Evaluate data availability, data requirements (do these match?)
- Specify the modelling methods and choose the tools
- Build (identify) the model:
    - Choose variables that reflect the physical processes
    - Collect, analyse and prepare the data
    - Build the model
    - Choose objective function for model performance evaluation
    - Calibrate (identify, estimate) the model parameters: if possible, maximize model performance by comparing the model output to past measured data and adjusting parameters
- Evaluate the model:
    - Evaluate *the **model uncertainty** related to different sources of uncertainty (data, parameters, model structure), model sensitivity*
    - Test (validate) the model using the "unseen" measured data
- Apply the model (and possibly assimilate real-time data)
- Evaluate results, refine the model

1.3 Sensitivity analysis

Sensitivity analysis (SA) is aimed at the discovery of which parameters or inputs have a substantial influence on the outputs (i.e. which ones are "important"), and *ranking* the parameters accordingly. SA method analyse how much changes in inputs or parameters influence the model output. It can be seen as a simplified and often fast way to assess how much the uncertainty in parameters or inputs propagates to the model output, i.e. allows for judging about the model uncertainty.

Local sensitivity analysis is the estimation of how sensitive the model output $y = M(x, \theta)$ to changes in parameters θ. The "sensitivity" of a model response y to a parameter θ is defined as the rate of change (slope) $\partial y/\partial\theta$ of the response y in the direction of increasing values of the parameter θ. A (simplified) traditional procedure for this is as follows:

- fix parameters θ_i at some reference value θ_i^* (typically, the calibrated one)
- perturb each of the parameters θ_i by a small amount $\Delta\theta_i$, *one at a time*
- compute the corresponding change Δy in the model output y
- compute the Sensitivity coefficients that reflect the slope of the input–output relationship at the reference point:

$$\Delta y \ / \ \Delta\theta_i$$

Weakness of this approach is that it provides only a limited view of model sensitivity because the results can vary with location in the factor space. such analysis is valid only locally, around $\{\theta_i^*\}$.

Global sensitivity analysis (GSA) provides more general results by characterizing the model sensitivity over the entire parameter space. GSA methods compute their indicators of sensitivity computed at a number of different points sampled across the domain of interest (i.e., a population sample), where the sample locations are selected (in some way) to be representative of the entire domain. One of the widely used methods is that the Morris One-

At-a-Time (MOAT) (Morris, 1991). Theoretic basis of this method is that the overall effect and interaction effect of each parameter can be approximated by the averaged mean and standard deviation of the gradients of each parameter calculated at points sampled in the parameter space.

Yet another way to carry out the GSA is to employ the *stepwise linear regression* analysis when a sequence of regression models linking the relevant variables is constructed (Draper & Smith 1981):

- start with the input variable that explains the largest amount of variance in the output;
- at each successive step, the input variable that explains the largest fraction of residual variance is included in the model;
- continue until all input variables that explain statistically significant amounts of variance in the output have been included in the model.

Weakness of this approach is that it can be typically applied to monotonic functions only.

Mutual information analysis shows how much information about y one obtains if x_i is known, and in this sense can be also used for global sensitivity analysis (x_i is a model input, and y is the model output). Average mutual information (AMI) is defined as follows:

$$\text{AMI}=\sum_{i,j} P_{XY}(x_i, y_j)\log_2\left[\frac{P_{XY}(x_i, y_j)}{P_X(x_i)P_Y(y_j)}\right]$$

where $P(x,y)$ is the joint probability for realisation x of X and y of Y; and $P(x)$ and $P(y)$ are the individual probabilities of these realisations. If X is completely independent of Y then AMI $I(X;Y)$ is zero. Of course, traditional correlation analysis can be used for the same purpose as well.

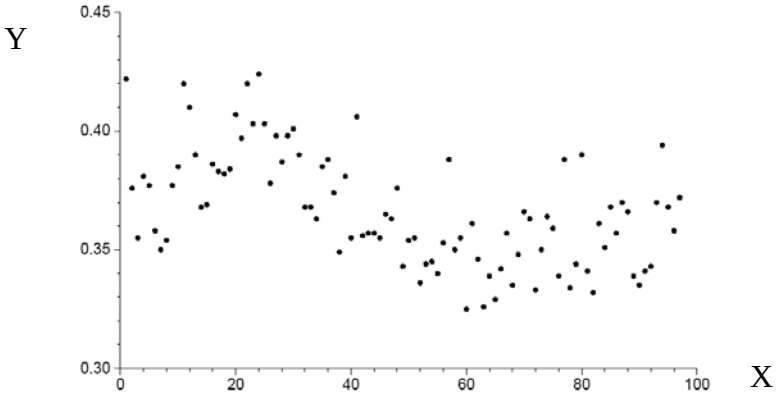
For a comprehensive coverage of various sensitivity analysis methods the reader is referred to the book by Saltelli et al. (2008) and the papers by Gan et al. (2014) and Song et al. (2015).

2 Some relevant notions of statistical modelling

Uncertainty is very often characterised using probability theory, statistics and statistical modelling. This chapter reminds of some relevant basic notions. For a good introductory text the reader is directed to the article in Wikipedia on statistics, and for the extensive coverage of the topic to many books on statistics and statistical model, e.g. Freedman (2005) and Larsen and Marx (2012).

2.1.1 Distributions

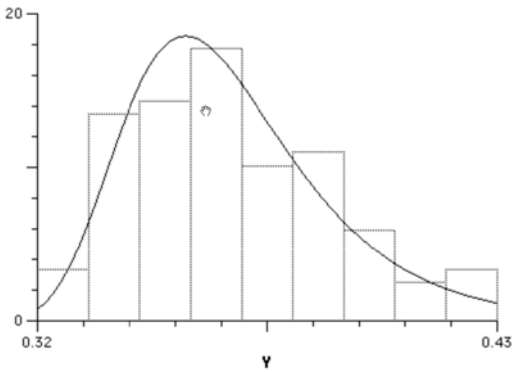
Consider an example (McLaughlin 1999). Data about two related variables X and Y is collected.



There are 97 observations of X and Y – data set D. There is some underlying model linking X and Y but at this stage we are interested only in the statistics of the output Y.

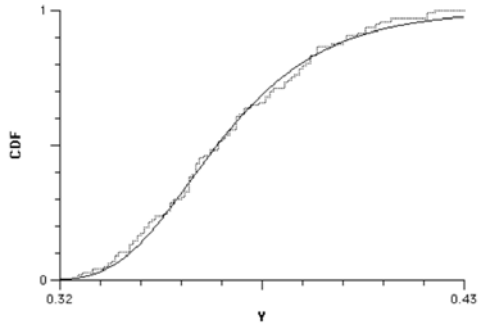
It is possible to build a histogram (bin is 0.011 wide) of 97 values of y. Then the probability density function (pdf) *f*

$$\text{Prob}\left(a \le y \le b\right) = \int_a^b f\left(y\right) dy$$



and cumulative distribution function CDF

$$\text{Prob}\left(y \le b\right) = F\left(b\right) = \int_{-\infty}^b f\left(y\right) dy$$



Pdf *f*(y) (solid curved line above the histogram) can be seen as a model describing the data about y; it is typically unknown and we want to estimate it from data. In statistical learning this is so-called model estimation problem. K% *quantile* is the value of *y* corresponding to the probability K.

After some experiments (possibly using software that automates the process) we may conclude that observed data can be described by the Gumbel distribution:

$$\mathrm{PDF} = \frac{1}{\mathrm{B}} \exp\left(\frac{\mathrm{A}-\mathrm{y}}{\mathrm{B}}\right) \exp\left(-\exp\left(\frac{\mathrm{A}-\mathrm{y}}{\mathrm{B}}\right)\right)$$

This model has two parameters A, B that are to be identified by maximizing the likelihood of D.

2.1.2 Likelihood

Let the observed data D (i.e. output var y) be described by some distribution *f* (y).

Likelihood of a dataset shows how "likely" this set of observations is if sampled by given pdf. Likelihood is the probability of observing data D. So for the independently sampled N values

$$\text{Likelihood} \equiv \prod_{k=1}^{N} \mathrm{f}\left(\mathrm{y}_k\right)$$

We assume data is generated by pdf *f*(y), but what are the values of A and B? We would like to choose them so, that the probability to observe data D, or in other words, the likelihood would be highest. The *maximum likelihood (ML)* is achieved if we choose parameters A=0.3555, and B = 0.01984. Any other set of parameters would make the observed dataset less likely to have been observed, so has to be rejected. Since D was indeed observed, these A, B are optimal.

2.1.3 Bias and variance, accuracy and precision

In data driven modelling the sources of uncertainty are similar to those for other hydrological models, but there is an additional focus on data partitioning used for model training and verification. Often data is split in a non-optimal way. A standard procedure for evaluating the performance of a model would be to split the data into training set, cross-validation set and test set. This approach is however, very sensitive to the specific sample splitting (LeBaron and Weigend, 1994). In principle, all these splitting data sets should

have identical or similar distributions, but we do not know the true distribution. This may bring additional uncertainty as well.

The prediction error of any regression model can be decomposed into the following three sources (Geman et al., 1992): (a) model bias; (b) model variance; and (c) noise. *Model bias* and *variance* may be further decomposed into contributions from data and contributions from the training process. Furthermore, noise can be also decomposed into target noise and input noise. Estimating these components of prediction error (which is however not always possible) helps to compute the predictive uncertainty.

The terms bias and variance come from a well-known decomposition of prediction error. Given N data points and M models, the decomposition is based on the following equality:

$$\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(t_i - y_{ij})^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - \bar{y}_i)^2 + \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\bar{y}_i - y_{ij})^2 \qquad (1)$$

where, t_i denotes the ith target, y_{ij} denotes the ith output of the jth model, and $\bar{y}_i = \frac{1}{M}\sum_{j=1}^{M} y_{ij}$ denotes the average model output calculated for input i.

The left-hand-side term of Eq. (4) is the well known mean squared error. The first term in the right hand site is the square of bias and the last term is the variance. The bias of the prediction errors measures the tendency of over or under prediction by a model, and is the difference between the target value and the model output. From Equation (4) it can be seen that the variance does not depend on the target, and measures the variability in the predictions by different models.

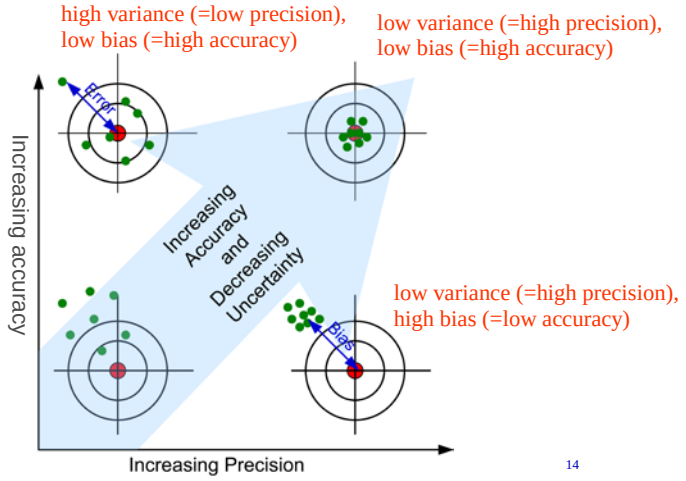


Figure 5. Illustration of the concepts: of Accuracy vs Precision, and Bias vs Variance

3 Main notions and sources of uncertainty

3.1 Notion of uncertainty

A common definition of something *uncertain* given for example by the Webster dictionary follows: not surely or certainly known, questionable, not sure or certain in knowledge, doubtful, not definite or determined, vague, liable to vary or change, not steady or constant, varying. The noun uncertainty results from the above concepts and can be summarizes as the state of being uncertain. However, in the context of modelling, uncertainty has a specific meaning, and it seems that there is no consensus about the very term of uncertainty, which is conceived with differing degrees of generality (Kundzewicz, 1995).

Of course uncertainty can be defined with respect to *certainty*. For example, Zimmermann (1997) defined certainty as: "certainty implies that a person has quantitatively and qualitatively the appropriate information to describe, prescribe or predict deterministically and numerically a system, its behaviour or other phenomena". Situations that are not described by the above definition shall be called uncertainty. Similar definition are given by Gouldby and Samuels (2005): "a general concept that reflects our lack of sureness about someone or something, ranging from just short of complete sureness to an almost complete lack of conviction about an outcome".

In the context of modelling, uncertainty can be defined as a state that reflects our lack of sureness about the outcome of a physical processes or system of interest and gives rise to potential difference between assessment of the outcome and its "true" value. More precisely, uncertainty of a model output is the state or condition that the output can not be assessed uniquely.

3.2 Two conceptual types of uncertainty

The two conceptual types of uncertainty are distinguished:

- *aleatoric* – associated with the randomness observed in nature, regarded as irreducible, and typically characterized by probabilities (pdfs). Also referred to as variability, irreducible uncertainty, and stochastic uncertainty, random uncertainty;
- *epistemic* – associated with the lack of knowledge, imprecision in data and observations, or linguistic ambiguity in expert knowledge. Also referred to as subjective uncertainty, and reducible uncertainty. Major paradigms to characterise these are interval mathematics, fuzzy set theory, rough sets theory, and Dempster-Shafer evidence theory, however, probability theory, due to its wide spread is used widely as well. Uncertainty resulting from insufficient information may be reduced if more information is available.

Often it is difficult to determine if uncertainty is aleatoric or epistemic, and as we collect more data and understand the physical processes better and better (i.e. increase our knowledge) the epistemic uncertainty is reduced but still it is difficult to know to what extent the remaining uncertainty is epistemic or can be described as fully aleatoric.

3.3 Uncertainty representation

Representation of uncertainty depends on how we treat it – as aleatoric (randomness) or as epistemic (lack of knowledge or imprecision in data).

3.3.1 Probabilistic representation

In most reported applications uncertainty was treated as aleatoric, and *probability theory* has been the primary tool for representing uncertainty in mathematical models. Different methods can be used to describe the degree of uncertainty. The most widely adopted methods use Probability Density Functions (pdf) of the quantity, subject to the uncertainty. However, in many practical problems the exact form of this probability function cannot be derived or found precisely.

If it is difficult to derive pdf it may be still possible to quantify the level of uncertainty by the calculated statistical moments such as the *variance, standard deviation,* and *coefficient of variation.* Another measure of the uncertainty of a quantity relates to the possibility to express it in terms of the two quantiles (e.g., 5% and 95%), or *prediction interval*. The prediction intervals consist of the upper and lower limits between which a future uncertain value of the quantity is expected to lie with a prescribed probability. The endpoints of a prediction interval are known as *the prediction limits*. The width of the prediction interval gives us some idea about how uncertain we are about the uncertain entity.

Among the UA probability-based methods the following can be mentioned: Monte Carlo, GLUE (Beven and Binley 1992), DREAM (Vrugt et al., 2009). Modelling frameworks explicitly including probabilistic uncertainty analysis were described by Wagener et al., 2001; Gupta et al, 2008; Refsgaard et al., 2007.

3.3.2 Fuzzy and other representations

Although useful and successful in many applications, probability theory is, in fact, appropriate for dealing with only a very special type of uncertainty, namely aleatoric (random) (Klir and Folger, 1988). For epistemic uncertainty corresponding to lack of data, vagueness or imprecision other methods and cannot be treated with probabilistic approaches. Major paradigms to characterise these are interval mathematics, fuzzy set theory (Zadeh, 1965, 1978) and *fuzzy measures*, rough sets theory, and Dempster-Shafer evidence theory (DST) (Shafer 1976).

There are also examples of this approach, especially with the use of fuzzy logic, however, it is much less popular than the probabilistic one. Some works on using fuzzy logic in characterization of uncertainty follow: Bardossy and Duckstein, 1995; Schulz and Huwe, 1997, 1999; Abebe et al., 2000; Maskey et al., 2004; Jacquin and Shamseldin, 2007; Jacquin, 2010). Much less has been done with DST (see however Raje and Mujumdar, 2010).

There are methods emerging that allow for combination of both types of uncertainty, for example, *fuzzy random variable (FRV) theory* that uses an extension of the classical random variable whose values are fuzzy numbers. An example: a variable takes random values around the mean, but the mean is a fuzzy number.

Another theory to consider is the "cloud theory" attributed to Deyi Li and colleagues (Li et al 2009, and references therein), where a membership function is not fixed but is represented by a "cloud" (in this respect it is conceptually close to Type-2 fuzzy logic by Jerry Mendel).

Information theory is also used for representing uncertainty. *Shannon's entropy* (Shannon, 1948) is a measure of uncertainty and information formulated in terms of probability theory.

In modelling the primary tool for handling uncertainty is still probability theory, and, to some extent, fuzzy logic. In practice same set of methods are often used to quantify both types of epistemic and aleatoric uncertainties. Indeed it can be justified by the fact that we do not know if a particular variable or factor is random or not, or there is a combination of both. An example: uncertainty in model parameters would be properly characterise as the lack of knowledge, i.e. as aleatoric, however most authors use probabilistic representation assuming thus epistemic uncertainty. A wider use of probability theory can be explained by the fact that it is a widely taught and known much more than other methods, and that probabilistic measures and approaches to design are widely adopted in civil engineering standards.

3.4 Sources of model uncertainty

Uncertainties affecting the model results stem from a variety of sources and are related to our understanding and measurement capabilities regarding the real-world system under study. The following sources are typically mentioned (e.g. Melching 1995; Gupta et al. 2005).

1. Perceptual model uncertainty, i.e. the conceptual representation of a system that is subsequently translated into mathematical form in the model. The perceptual model (Beven 2001) is based on our understanding of the real-world system, i.e. in hydrology these would be flowpaths, number and location of state variables, runoff production mechanisms, etc. This understanding might be poor and therefore our perceptual model might be highly uncertain (Neuman 2003).

2. Data uncertainty, i.e. uncertainty caused by errors in the measurement of input and output data, or by data processing. Additional uncertainty is introduced if long-term predictions are made, for instance in the case of climate change scenarios for which no observations are available. A water motion model might also be applied in integrated systems, e.g. connected to a socio-economic model, to assess for examples impacts of water resources changes on economic behavior. Data to constrain these integrated models is rarely available (e.g. Letcher et al. 2004). An element of data processing uncertainty is introduced when a model is required to interpret the actual measurement. For example, a weather radar is used to measure reflectivity that have to be transformed to rainfall estimates using a (empirical) model with a chosen functional relationship and calibrated parameters, both of which can be highly uncertain. Yet another type of uncertainty is spatial uncertainty due to inaccuracies of representation of spatial elements in GIS (Brown and Heuvelink 2007) and especially in DEM (Moya Quiroga et al., 2011); these could important to take into account in flood inundation studies.

3. Parameter estimation uncertainty, i.e. the inability to accurately identify the model parameters values based on the available information. Parameters can be estimated either by direct measurement, expert assessment, or as a result of model calibration that in principle should lead to a "best possible vector" of parameters. However all these methods

carry a lot of uncertainty with them: measurements are never perfect, expert judgment is often subjective, and process of calibration involves the measured (past) model outputs that are inaccurate as well.

4. Model structural uncertainty brought by inevitable simplifications and ambiguity in the description of real-world processes. Often there is also some initial uncertainty in the model state(s) at the beginning of the modeled time period requiring a model "warm-up". All this leads to inaccurate model predictions.

3.5 Relationship between model complexity and model uncertainty

One may expect that increasing a model complexity (i.e. making it more adequate) should lead to reduction in output uncertainty. However, paradoxically, in reality situation is much more complex. Figure 2 presents how different sources of the uncertainty might vary with model complexity. As the model complexity (and the detailed representation of the physical process) increases, structural uncertainty decreases. However, with the increasing complexity of a model, the number of inputs and parameters also increases and consequently there is a good chance that input and parameter uncertainty will increase. Due to the inherent trade-off between model structure uncertainty and input/parameter uncertainty, for every model there is the optimal level of model complexity where the total uncertainty is minimum.

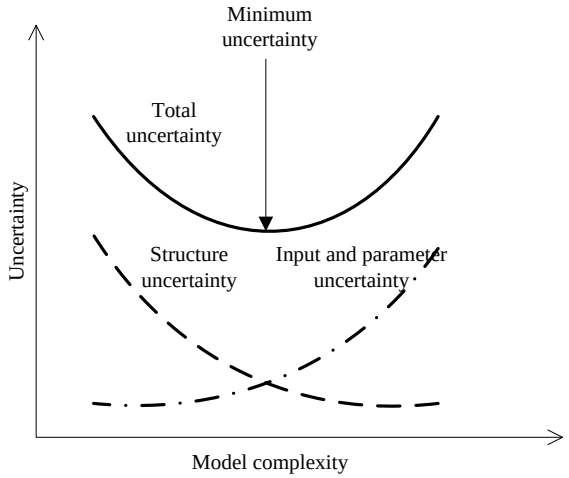


Figure 6. Dependency of various sources of uncertainty on the model complexity.

3.6 Propagation of probabilistic uncertainty through models

The central step in model uncertainty analysis is studying how uncertainty propagates through the model. Consider a model M that attempts to reproduce the measured value y, such that

$$y = M(x, \theta) + \varepsilon \quad (2)$$

where y = measured output, x = input, θ = parameters, ε = model error. The model error ε is a manifestation of all sources of uncertainty that propagate through the model to its output. The output of model M is usually denoted as y^ so that

$$\hat{y} = M(x, \theta) \quad (3)$$

Figure 3 shows schematically a possible relationship between the actual value of output y^a, the measured value y and the model output y^. Obviously, since the actual value is not known, model error is the difference between the observed value (that represents the actual value with some error) and the modelled value.

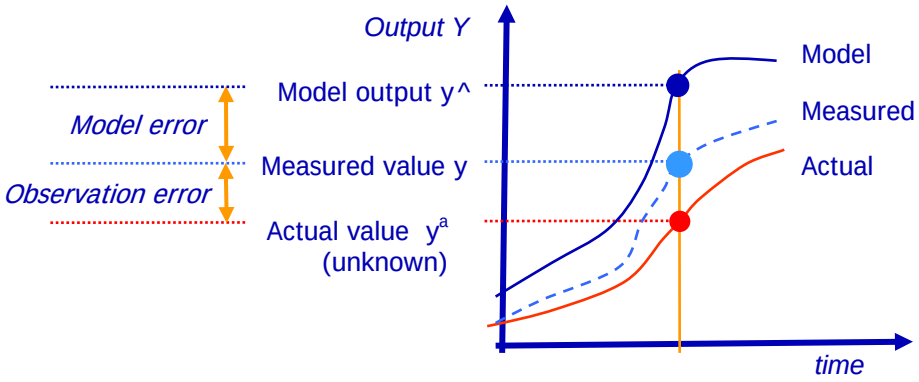


Figure 7. Model error in reproducing the measured value y

As discussed earlier, there are many sources of uncertainty and the error ε may have many components, so in more detail the model equation can be presented as follows:

$$y = M(x, \theta) + \varepsilon_s + \varepsilon_\theta + \varepsilon_x + \varepsilon_y \quad (4)$$

Schematically, propagation of uncertainty in X and θ to output y can be depicted as follows:

pdf of parameters → pdf of output or pdf_θ → pdf_y
pdf of inputs pdf_x → pdf of output or pdf_x → pdf_y

In this equation the following groups (types) of uncertainty of y are present:

1) Observational (data) uncertainty (ε_x , ε_y)
   a) measurement errors due to both instrumental and human errors
   b) error due to inadequate representativeness of a data sample due to scale incompatibility or difference in time and space between the variable measured by the instrument and the corresponding model variable.

2) Model (structural) uncertainty ε_s which is associated with inadequacies of the model structure.

3) Parametric uncertainty ε_θ which is associated with the lack of information or inaccurate measurements of the model parameters.

3.7 Main approaches and steps in uncertainty analysis

3.7.1 General

Once the uncertainty in a model is acknowledged, it should be analyzed and quantified with the ultimate aim to reduce the impact of uncertainty. There is a large number of uncertainty analysis methods published in the academic literature. Pappenberger et al. (2006) provide a

decision tree to help in choosing an appropriate method for a given situation. Uncertainty analysis process varies mainly in the following: (a) type of models used, (b) source of uncertainty to be treated; (c) the representation of uncertainty (probabilistic or fuzzy); (d) purpose of the uncertainty analysis; (e) availability of resources.

Uncertainty analysis has comparatively long history in physically based and conceptual modelling (see, e.g., Beven and Binley, 1992; (Kuczera and Parent (1998); Gupta et al., 2005; Melching (1995); Montanari (2007); Shrestha and Solomatine (2008)).

3.7.2 Main approaches to analyse residual and data uncertainty

It is reasonable to classify the main approaches to uncertainty analysis with respect to the two main types of model uncertainty that can be distinguished:

A. The *residual uncertainty* of models. In this case the model parameters and/or model inputs are considered to be fixed (deterministic), i.e. the model is considered to be optimal (calibrated) and deterministic. Model error is considered as the manifestation of uncertainty. If there is enough past data about the model errors (i.e. it uncertainty), it is possible to build a statistical or machine learning model of uncertainty trained on this data.

Two methods can be mentioned: (a) quantile regression (QR) method (Koenker and Basset, 1978; Koenker, 2005) in which linear regression is used to build predictive models for distribution quantiles, and (b) a more recent approach that takes into account the input variables influencing such uncertainty and uses more advanced machine learning methods (neural networks, model trees etc.) – the UNEEC method (Shrestha and Solomatine 2008; Solomatine and Shrestha 2009).

Quantile regression (QR) introduced by Koenker and Bassett (1978) is a statistical technique intended to estimate the conditional quantile functions. Just as the classical linear regression methods based on minimizing sums of squared residuals enable one to estimate the models for conditional mean of the response variable, given certain values of the predictor variables, QR methods offer a mechanism for building the models for the conditional median function, and the full range of other conditional quantile functions. Note that in opposition to UNEEC, QR is a linear regression method and is based on the whole data set, and does not include the local specialized nonlinear models as is done in UNEEC.

B. The *data uncertainty* (parametric and/or input) – in this case we study the propagation of uncertainty (presented typically probabilistically) from parameters or inputs to the model outputs.

In case of simple functions representing models analytical approaches can be used (see, e.g., Tung, 1996), or approximation methods (e.g., first-order second moment method (Melching, 1992)). However, for real complex non-linear models implemented in software there is no other choice except using some variant of the *Monte Carlo simulation* when values of parameters or inputs are sampled from the assumed distributions and the model is run multiple times to generate multiple outputs (Kuczera and Parent, 1998; Beven and Binley, 1992). This is the most widely used approach, and it will be considered in this text.

With this in mind, one may consider the approach to stepwise “building up” the model uncertainty as follows:

- first consider the *residual uncertainty* of an optimal model M (X, *p**)
- then add and consider the model uncertainty due the *parameters uncertainty* (p)
- then add and consider the model uncertainty due the *data* (mainly, *input*) *uncertainty* (X)
- then add and consider the *structural uncertainty* of the model M (X, *p*).

All types of uncertainty, being considered together, provide the uncertainty of a model class M(X, *p*), given the probabilistic properties of parameters and data.

3.7.3 Main steps in uncertainty analysis

The following steps of UA are typically considered:

- Identification of sources of uncertainty (Input, parameter, model, operational)
- Quantification of uncertainty, i.e. prior (assumed) distribution and range of parameters
- Reduction of uncertainty (e.g. better understanding of the model, accurate measurement etc)
- Studying *propagation of uncertainty through the model* (e.g. by Monte Carlo simulation)
- Quantification of uncertainty in the model outputs, e.g. analysis of the outputs of the results)
- Application of the uncertain information in decision making process

4 Monte Carlo simulation approaches

4.1 General Monte Carlo framework

Analytical and approximation methods can hardly be applicable in case of using complex computer-based models. Here we will present only the most widely used probabilistic methods based on random sampling – Monte Carlo simulation method. MC simulation is a flexible and robust method capable of solving a great variety of problems. In fact, it may be the only method that can estimate the complete probability distribution of the model output for cases with highly non-linear and or complex system relationship (Melching, 1995). It has been used extensively and as a standard means of comparison against other methods for uncertainty assessment. In MC simulation, random values of each of uncertain variables are generated according to their respective probability distributions and the model is run for each of realizations of uncertain variables. Since we have multiple realizations of outputs from the model, standard statistical technique can be used to estimate the statistical properties (mean, standard deviation etc.) and empirical probability distribution of the model output. MC simulation method involves the following steps:

1. Randomly sample uncertain variables X_i from their (joint) probability distributions. These variables are typically model parameters or/and model inputs.
2. Run the model $y = M(x_i)$ with the set of random variables x_i.
3. Repeat the steps 1 and 2 s times, storing the realizations of the outputs $y_1, y_2,...,y_s$.
4. From the realizations $y_1, y_2,\ldots, y_s$, derive the cdf and other statistical properties (e.g. mean and standard deviation) of Y.

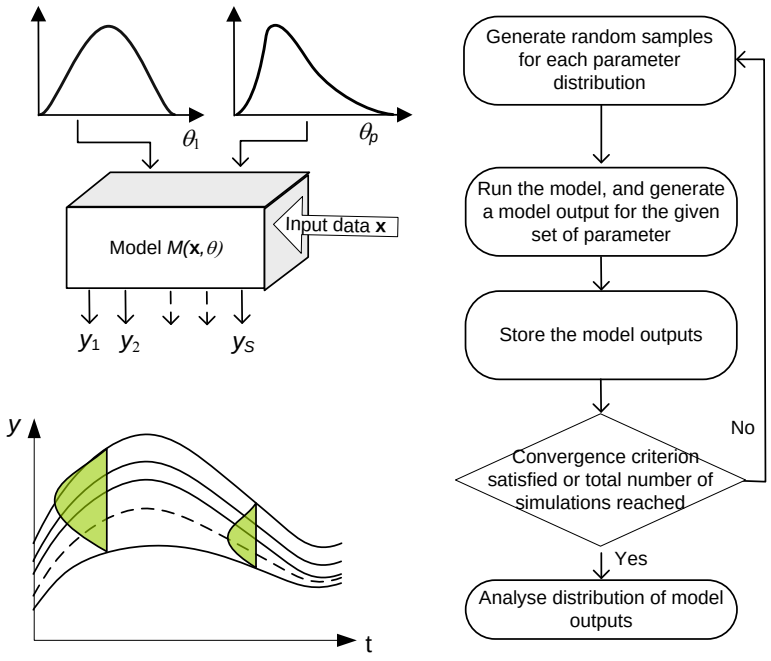


Figure 8. Monte Carlo simulation process for the analysis of parametric uncertainty. Parameter vectors θ are sampled from the assumed (non-joint) distributions; the inputs **x** are considered to be certain and are not sampled. The model M is run for each vector and multiple output time series $y(t)$ are generated. Uncertainty of outputs y is represented by the resulting empirical *pdf* (green shapes) for each time step

Figure 9 presents using MC in investigation of parameter uncertainty of the model M (x, θ). Parameter vectors θ are sampled from the assumed distributions, the model is run for each vector and multiple output time series $y(t)$ are generated. Uncertainty of outputs y is represented by the resulting empirical *pdf* (or its characteristics, like mean, variance, or the 5% and 95% quantiles) for each time step.

When MC sampling is used, the error in estimating pdf is inversely proportional to the square root of the number of runs s and, therefore, decreases gradually with s. As such, the method is computationally expensive, but can reach an arbitrarily level of accuracy. The MC method is generic, invokes fewer assumptions, and requires less user-input than other uncertainty analysis methods.

4.2 Details of the Monte Carlo approach

In it simplest form, MC process involves random sampling of parameter vector (typically, for all parameters independently) from distributions, typically with independent uniform or normal distributions for each parameter. The model is then run with many such parameter sets, producing multiple sets of model output. These are used together to generate uncertainty intervals for model predictions.

It should be mentioned however, that the MC method suffers from a number of practical limitations, the two most important being: (i) it is difficult to sample the uncertain variables from their joint distribution unless the distribution is well approximated by a multinormal distribution (Kuczera and Parent, 1998); and (ii) it is computationally expensive for complex models.

There have been methods developed that give this process more structure, or introduce variations in sampling or the ways to combine the ensembles of the generated outputs. Various variants of this approach differ in the following:

4.2.1 Sampling strategy

In low-dimensional problems (small number of parameters, e.g. 2-3) and simple fast-running models it is possible to sample sufficiently many parameter vectors and to cover the parameter space reasonably well (i.e. with high density). In such "data-rich" cases it may be relatively easy to estimate the summary statistics of the resulting distribution (mean, standard deviation etc.). However, in high-dimensional problems and slow-running models the number of sampled points would be lower, and they cannot cover the whole (highly-dimensional) parameter space well. In this case usage of an "economical" sampling strategy becomes of great importance. The following approaches can be considered:

a) *Pure random sampling*. all vectors sampled randomly and independently of each other and the models runs are independent; this approach can be used for low-dimensional cases.

b) *Latin Hypercube sampling* introduced (McKay et al. 1979) – see Figure 6. In it, the range of each variable is divided into M equally probable intervals, and M sample points are placed as shown the number of divisions, M, is equal for each variable. Important feature of LHS is that it does not require more samples for more dimensions (variables).

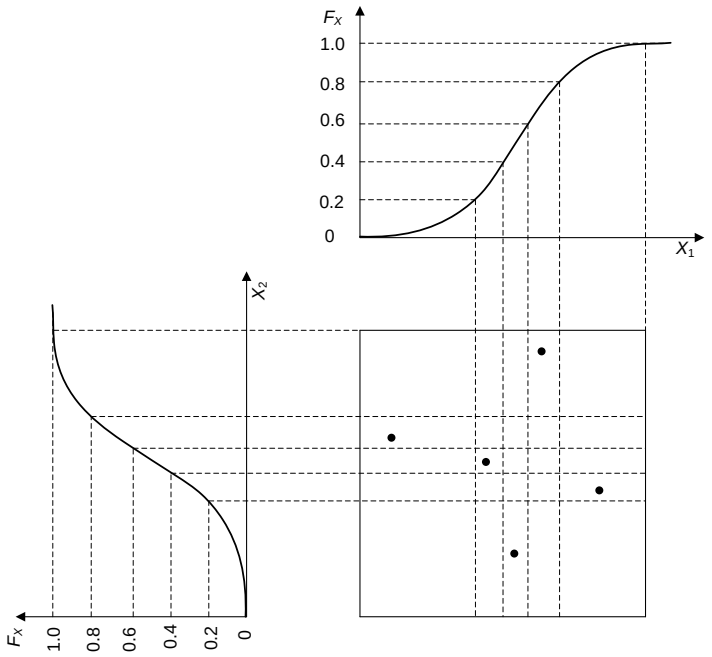


Figure 10. Latin Hypercube Sampling (LHS).

c) *Consecutive sampling*. Each consecutive vector depends on the results of the previously generated vectors and on the corresponding model runs. This is an approach taken in the Monte Carlo Markov Chain (MCMC) methods; the most widely used version of this approach is the Metropolis-Hastings (MH) algorithm, a variation of the Metropolis algorithm of Metropolis et al. (1953) that was proposed by Hastings (1970). Other methods, initially developed for randomized search, and by design covering the most important areas of the parameter space, can be used as well, e.g. the ACCO algorithm (Solomatine, 1999). There are sampling strategies that combine both approaches, e.g. SCEM-UA algorithm (Vrugt et al., 2003) (combination of the SCE-UA randomized search algorithm combined with the Metropolis algorithm), or DREAM algorithm (Vrugt et al., 2009) (combination of differential evolution optimization algorithm with MCMC scheme).

4.2.2 Stopping rules

There could various stopping rules employed:

- MC simulations are run for a fixed number of times;
- MC simulations are run till some stopping criteria are met, for example, convergence of variation in the average model error across all model runs.

4.2.3 Formulation of likelihood

To estimate the model performance, the notion of likelihood is used. Classical formulation defines it as the probability of observing the sampled data set if a certain sampling distribution is used. Good models have high likelihood. However there are also alternative formulations of likelihood that may be formally quite far from being probabilities. In this respect the following division of methods is possible:

- Classical formulation of likelihood
- Alternative (empirical) formulations, like in the Generalized Likelihood Uncertainty Estimation (GLUE) approach (Beven and Binley (1992))

4.2.4 Discarding (or not) some models based on their performance

On the basis of their performance, some models may be considered as "bad" ("non-behavioral" in terminology of Beven and Binley (1992)), if the corresponding value of likelihood is below a certain threshold (cutoff value). In this case these models' outputs are excluded from being part of the ensemble of outputs used to estimate the model uncertainty. There are methods that

- Explicitly employ the cutoff threshold, e.g. the Generlaized Sensitivity Analysis (GSA) (Spear and Hornberger, 1980), and the **GLUE approach** (Beven and Binley (1992));
- Do not impose any rigid cutoff value for likelihood value and allow for all sampled parameter vectors (models) to be part of the ensemble. This is the case for most MC-like methods (like BaRE approach by Thiemann et al., 2001; Stedinger et al., 2008 and many others) that strictly follow the statistical learning and Bayesian theory.

A number of authors (Thiemann et al., 2001; Montanari, 2005; Mantovan and Todini, 2006; Stedinger et al., 2008) point at the fact that the popular GLUE methodology does not formally follow the Bayesian approach in estimating the posterior probabilities of parameters and of the output distribution. At the same time, Vrugt et al., 2008b presents examples when under a variety of different conditions both Bayesian and informal Bayesian methods (like GLUE) can result in similar estimates of predictive uncertainty.

4.3 Selecting the best model

Consider model $y^{\wedge} = M(x, \theta)$ and observations y. Model can be run at different time steps so it is assumed that y and $y^{\wedge}$ are in fact time series (vectors) representing data for several times steps $k = 1,2,...N$. Obviously we want error $\varepsilon = y^{\wedge} - y$ be at a minimum. This is done by calibration, i.e. choosing the appropriate parameter vector θ. Likelihood L is defined as *the probability that the sample y was observed, given the underlying model.* In other words, parameters θ of a model are the "right" ones, given the observed sample y. Obviously, we want to maximize this probability L.

Consider residuals ε_k (model errors) at time *k*. We assume they are independent and normally distributed with mean A and standard deviation B.

$$\varepsilon_k = \hat{y}_k - y_k, \quad are \sim Normal(A_k B_k)$$

Typically it is also assumed that errors are unbiased, i.e. mean A_k is zero. Then likelihood is given by multiplying N probabilities (one for each time step *k*):

$$Likelihood \equiv \prod_{k=1}^{N} f(\varepsilon_k) = \prod_{k=1}^{N} \left[\frac{1}{B_k \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\varepsilon_k}{B_k} \right)^2 \right) \right]$$

It is easier to work with the log-likelihood; reason is that we would not have a danger of having thousands of multiplications of very small numbers, and by taking a logarithm multiplications are converted to additions:

$$Log-likelihood = -\sum_{k=1}^{N} \log\left(B_k \sqrt{2\pi} \right) - \frac{1}{2} \sum_{k=1}^{N} \left(\frac{\varepsilon_k}{B_k} \right)^2$$

$$Maximum \quad likelihood \Rightarrow -\min\left[\sum_{k=1}^{N}\left(\frac{\varepsilon_k}{B_k}\right)^2\right]$$

If all points are from the same Normal distr., then B_k are all equal, and

$$Maximum \quad likelihood \Rightarrow \frac{1}{B}\min\left[\sum_{k=1}^{N}\varepsilon_k^2\right] \Rightarrow \min\left[\sum_{k=1}^{N}\left(\hat{y}_k - y_k\right)^2\right]$$

If all errors are ~Normal(0, B), then the parameters that minimize Squared Error are the very same parameters that maximize the likelihood of the residuals – i.e. *minimum Squared Error ensures maximum Likelihood*.

This in fact explains why most of the formulations of model error use squared errors. For example, a widely used root mean squared error, RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t))^2}$$

Of course, the minimum squared error ensures maximum likelihood only if assumption of normality and independence of model errors at each time step holds. These assumptions are never true but in many instances distribution of errors could be not too far from normal, so calibrating models on squared errors, like it is done in majority of cases, is not a bad idea. If there are good reasons not to believe these assumptions, and there are ways to estimate (joint) model error distributions (which is very rare), it would be methodologically right to develop another formulation of optimal estimation of model error (possibly different from the squared error) using an approach presented in the derivations of maximum likelihood above.

Note that in equifinality principle (Beven and Binley 1992) is based on an idea that it is impractical to assume the existence of a single model, and it is right to select a number of models to be used in an ensemble setting. However, in practice of water modelling, practitioners tend to use only one, calibrated, model.

4.4 Machine learning predictive models of parametric or input uncertainty

MC simulation provides an average estimate of a model uncertainty across the time period for which data is available and model runs are made.

There is however a problem: the estimates are averages across the whole domain of different conditions or situations that the model is describing (characterised, e.g. by hydro-meteorological variables like temperature, air pressure, rainfall, flow, etc. that can be input, state or output variables). However model uncertainty could be (and is) different for different conditions. A question arises here: is it possible to make more accurate characterisation of a model uncertainty specific to a particular situation, rather than using average estimates from MC simulations?

Such characterisation of model uncertainty can be called a model of model uncertainty. This model can be used to assess the uncertainty of a model for the future time moments when new input data is fed into the model.

This is the main idea of the MLUE approach (Machine Learning for Uncertainty Estimation). Consider model (1) but with the state variables explicitly mentioned:

$$y^{\wedge} = M(x, S, \theta) \tag{5}$$

We will consider the joint space {x, S, y} in which model operates, and a particular situation of model application at time *t* can be characterised by a set of variable values possibly taken at various moments of time (lagged), for example, $\{x_t, S_t, y_t\}$, or $\{ x_t, x_{t-1}, x_{t-2}, S_t, y_t\}$, or $\{x_{t-1}, x_{t-2}, , g(x_{t-3}, x_{t-4}, x_{t-5}), S_t, S_{t-1}, y_t, y_{t-1}, y_{t-2}\}$, etc., where *g* is some function used, for example for weighted averaging. Denote such vector of characteristic variables as {s}, its dimension as *m*, and its realization at time *t* as $\{s_t\}$. The model output $y_t^{\wedge}$ at time *t* would have a corresponding vector $\{s_t\}$ (if lagging allows).

The main steps of the MLUE approach are:

- Analyze the dependency of the uncertainty measures (quantiles) on various combinations of (lagged) {x, S, y}, using, for example, correlation and average mutual information analysis;
- Select the characteristic variables $\{X_u\}$ to be used as input variables for the machine learning model U;
- Train the machine learning model(s) U_i to predict the uncertainty measures (e.g., quantiles) of MC realizations $QUANT_i = U_i (X_u)$
- Validate machine learning model U by estimating the uncertainty measures with the "new" input data

5 An illustrative case study

In this section parametric uncertainty analysis of a hydrological model will be demonstrated. Two methods will be applied: 1) Monte Carlo simulation, and 2) the MLUE approach in which a machine learning (non-linear regression) model of the results of MC analysis is built, and used to predict the uncertainty of model outputs in the future. It is based on the paper by Shrestha et al., 2009.

5.1 Study area

The Brue catchment located in South West of England, UK is selected for the application of the methodology. The catchment has a drainage area of 135 km^2 with the average annual rainfall of 867 mm and the average river flow of 1.92 m^3/s, for the period from 1961 to 1990. The discharge is measured at Lovington. The hourly potential evapotranspiration was computed using the modified Penman method recommended by FAO (Allen *et al.*, 1998). Splitting of available data set is based on Shrestha and Solomatine (2008), one year hourly data from 1994/06/24 05:00 to 1995/06/24 04:00 was selected for calibration of HBV hydrological model and data from 1995/06/24 05:00 to 1996/05/31 13:00 was used for the verification (testing) of the hydrological model. Each of the two data sets represents almost a full year of observations, and it appeared that the statistical properties of these sets are similar.

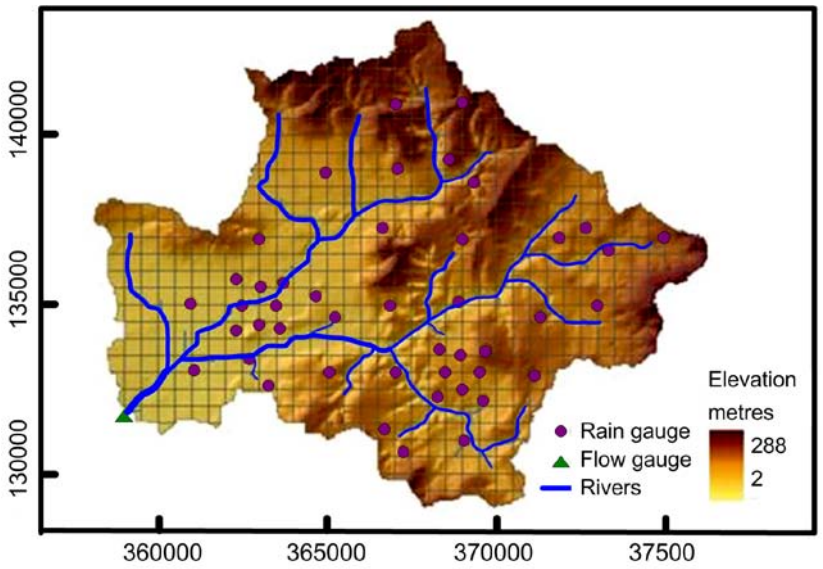


Figure 11: The Brue catchment showing (the horizontal and vertical axes refer to the easting and northing in British national grid reference co-ordinates).

5.2 Conceptual hydrological model

A simplified version of HBV model (Fig 2) was used to simulate river flows. Input data are observations of precipitation, air temperature, and estimates of potential evapotranspiration. The detailed description of the model can be found in Lindström *et al.* (1997).

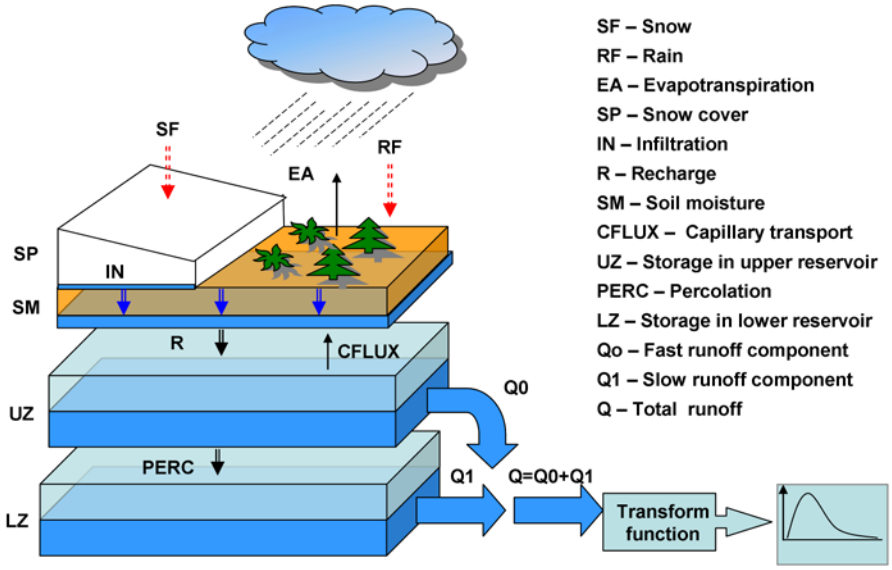


Figure 12: Schematic representation of HBV-96 model (Lindström et al., 1997) with routine for snow (upper), soil (middle) and response (bottom).

5.3 Experimental Setup

The nine parameters listed on Table 1 are used in HBV model. The model was first calibrated using the global optimization routine - adaptive cluster covering algorithm, ACCO (Solomatine *et al.*, 1999) to find the best set of parameters, and then followed by manual adjustments of the parameters. The ranges of parameters for automatic calibration and parameter uncertainty analysis were set based on a range of calibrated values from other model applications (e.g., Braun and Renner, 1992) and the information about the catchment.

Table 1: The prior intervals of the HBV model parameters used for calibration of the model and for parameter uncertainty analysis by MC method

Parameter	Description	Uniform prior range	Optimum value
FC	Maximum soil moisture content	100 - 300	160.33
LP	Limit for potential evapotranspiration	0.5 - 0.99	0.527
ALFA	Response box parameter	0 - 4	1.54
BETA	Exponential parameter in soil routine	0.9 - 2	1.963
K	Recession coefficient for upper tank	0.0005 - 0.1	0.001
K4	Recession coefficient for lower tank	0.0001 - 0.005	0.004
PERC	Percolation from upper to lower response box	0.01 - 0.09	0.089
CFLUX	Maximum value of capillary flow	0.01 - 0.05	0.038
MAXBAS	Transfer function parameter	8 - 15	12.00

The model was calibrated using Nash-Sutcliffe coefficient of efficiency (*CE*) as a performance measure of HBV model. For the calibration period *CE* was 0.96. The model was validated by simulating the flows for the independent validation data set, and *CE* was 0.83. Figure 9 shows the simulated hydrograph in both calibration and verification period.

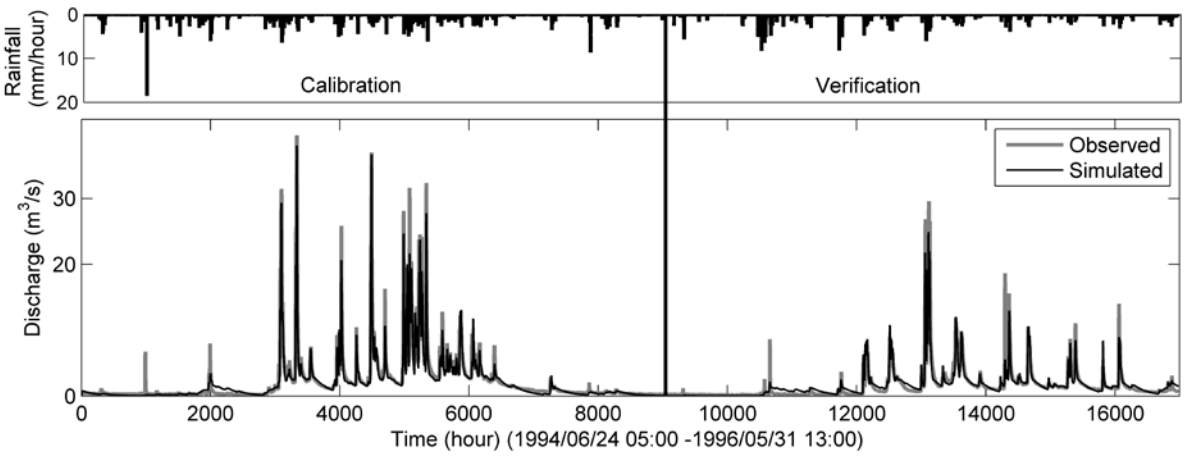


Figure 13: Observed and simulated discharge hydrograph

5.4 MC simulation and its convergence analysis

The parameters of HBV model are sampled from the uniform distribution with the ranges given in Table 1. The model is run for each random parameter set and likelihood measure is computed for each model run. *CE* is used as the basis to calculate the likelihood. We investigated the number of behavioral samples retained out of 74,467 MC samples for different values of rejection threshold. It was observed that only 1/3 of simulations (25000 samples) are accepted for threshold value of 0, whereas less than 1/10 of simulations are retained for a threshold value of 0.7.

We have also tested the convergence of MC simulations to know the number of samples required to get the reliable results. Mean and standard deviation of *CE* were used to analyze the convergence of MC simulations (Equ 12 and 13). Other statistics for convergence test can be found in Ballio and Guadagnini (2004).

$$ME_k = \frac{1}{k}\sum_{i=1}^{k} CE_i \quad (6)$$

$$SDE_k = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(CE_i - ME_k)^2} \quad (7)$$

where CE_i is the coefficient of model efficiency of ith MC run, ME_k and SDE_k are the mean and standard deviation of efficiency of model runs upto kth runs, respectively.
The Figure 10 depicts the two statistics – mean and standard deviation of *CE* – used to analyse the convergence of MC simulations. It is observed that both statistics are stable after 5000-10000 simulations, so 10,000 MC simulations are reasonable to consider in this case study.

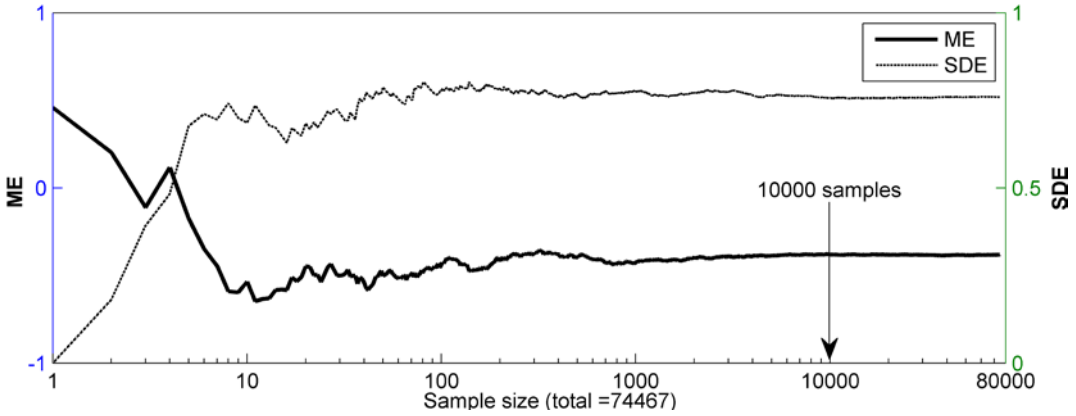


Figure 14: The convergence of mean and standard deviation of the coefficient of model efficiency. Note that x-axis is log scale to see the initial variation.

5.5 Sensitivity of parameters

The interaction between parameter values results in the broad region of acceptable simulations (corresponding to the different parameter values) and it is the combination of parameter values that produces the acceptable or non-acceptable simulations within the chosen model structure. Such result reveals little about the sensitivity of the model predictions to the individual parameters, except where some strong change in the likelihood measure is observed in a certain range of a particular parameter. More detailed analysis based on the extension of regional sensitivity analysis (Freer et al., 1996, Spear and Hornberger, 1980) was performed. Figure 11 shows the cumulative distributions for 10 equal sets (by number) of MC simulations. Each parameter population is ranked from best to worst in terms of the chosen likelihood function and the ranked population is then divided into ten bins of equal size. The cumulative likelihood distribution of each group is then plotted. Parameter sensitivity can be evaluated by assessing the spread of cumulative distribution function of each group. Parameters with the strong deviation of the distribution function are recognized as sensitive parameters. Three parameters *ALFA*, *K* and *MAXBAS* were found to be the most sensitive parameters, while the two parameters *K4* and *CLUX* shows no sensitive at all. The rest of the parameters show moderate sensitivity.

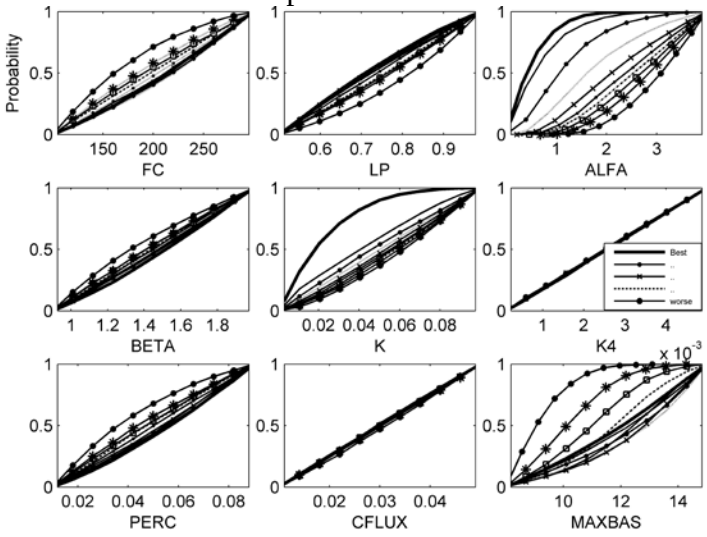


Figure 15: Sensitivity of individual parameters expressed as cumulative distribution values in 10 equal sets (by number) of MC simulations. Parameters with strong deviation of the distribution function are recognized as sensitive parameters (ALFA, K and MAXBAS).

5.6 Application of the MLUE method of uncertainty analysis

In the MLUE approach, a machine learning (non-linear regression) model of the results of MC analysis is built, and used to predict the uncertainty of model outputs in the future. See Shrestha et al., 2009 for details.

5.6.1 Artificial neural network (ANN) as an emulator of MC simulation

Once having the uncertainty results generated by MC simulation, ANN (which is, in fact, a non-linear regression model) is trained to learn functional relationship between the uncertainty results and the input data. The GLUE method has been used for parameter uncertainty estimation of HBV model. The threshold value of 0.0 (measure by *CE*) is selected to classify simulation either behavioural or non-behavioural. 90% uncertainty bounds are calculated using the 5% and 95% quantiles of the MC simulation realizations.

Corresponding 90% lower and upper PIs are calculated using the model output simulated by optimal model parameter sets found by Shrestha and Solomatine (2008) Hence the computed upper and lower PIs are conditional on contemporary value of the model simulation. Next step was to select the most relevant input variables to build predictive ANN model.

5.6.2 Selection of input variables for the ANN model

To select the input variables several approaches can be used (see, e.g., Solomatine and Dulal, 2003, Guyon and Elisseeff, 2003, Bowden et al., 2005). The input variables for model *V* are constructed from the rainfall, evapotranspiration and observed discharge. Experimental results show that evapotranspiration alone does not have significant influence on the prediction interval. Thus it was decided not to include the evapotranspiration as a separate variable, but rather to use effective rainfall (rainfall minus evapotranspiration for rainfall greater than evapotranspiration and zero otherwise).
Figure 12 shows the correlation coefficient and the average mutual information (AMI) of R_t and its lagged variables with the lower and upper PIs. It is observed that the correlation coefficient is minimum at the zero hour lag time and increases with the lags upto 9 hours (Figure 12a) for lower PI. While the optimal lag time (time at which the correlation coefficient and/or AMI is maximum) is 7 hours in the case of upper PI (Figure 12b). At this optimal lag time, information about the PIs is contained maximally in the variable R_t. Such findings are also supported by the AMI analysis. Additionally, correlation and AMI between the PIs and observed discharge were analysed. The results show that the immediate and the recent discharges (with the lag of 0, 1, 2) have very high correlation with the PIs. So it was also decided to use the observed discharge as the input to the model *V*.

Several structures of the input data including lagged variables were considered. The principle of parsimony was followed to avoid the use of a large number of inputs, so the aggregates, such as the moving averages or derivatives of the inputs that would have hydrological meaning were considered. Typically the rainfall depth at hourly time step partially exhibits random behaviour, and is not very representative of the rainfall phenomenon during a short period. Hence, we used the mean rainfall value of the lagged variables R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, and R_{t-9}, as the rainfall input, which is denoted as $R_{t\text{-}9a}$. Furthermore, derivative of flow indicates whether the flow situation is either normal or characterized by the base flow (zero or small derivative), or can be characterized as the rising limb of the flood event (high positive derivative), or the recession limb (high

negative derivative. Therefore, in addition to the flow variable Q_{t-1}, rate of change of flow at time t-1 is computed by deducting the flow at time t-2 from flow at t-1 and denoted by ΔQ_{t-1}..

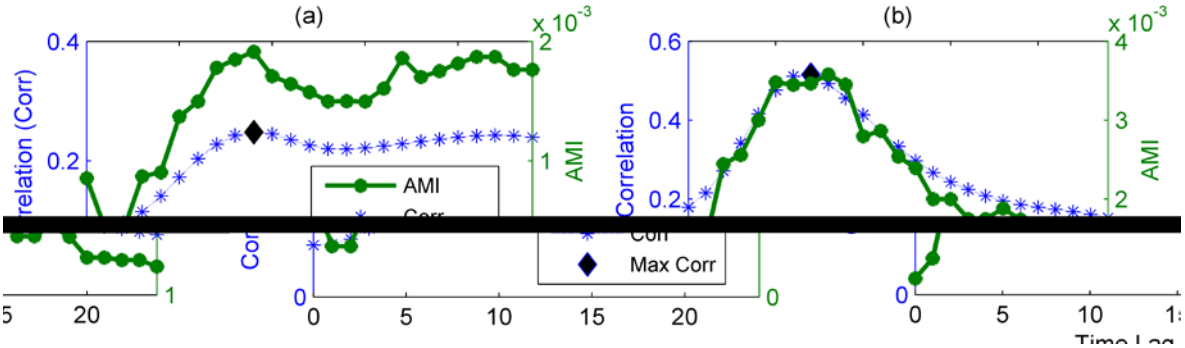


Figure 16: Linear correlation and AMI between rainfall R_t (a) lower PIs; and (b) upper PIs for different lags time.

The ANN model is based on 8745 data records from 1994/06/24 5:00 to 1995/06/24 4:00 (hourly) and its structure is given by:

$$PI = V(R_{t-9a}, Q_{t-1}, \Delta Q_{t-1}) \tag{8}$$

where PI = prediction intervals PI^L or PI^U
R_{t-9a} = moving average based on R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, and R_{t-9}
$\Delta Q_{t-1} = Q_{t-1} - Q_{t-2}$ (characterizes the change in the previous discharge).

5.6.3 Model training

The same data set used for calibration and verification of HBV model were used for training and verification of model V respectively. However, for proper training of the ANN model the calibration dataset is segmented into two sets, 15% of datasets for cross validation (CV) and as 85% for training.

CV data set was used to identify the best structure of ANN. In this paper, a multilayer perceptron network was used; optimization was performed by the Levenberg-Marquardt algorithm. The hyperbolic tangent function was used for the hidden layer with linear transfer function at the output layer. The maximum number of epoch was fixed to 1000. Trial and error method is adopted to detect the optimal number of neurons in the hidden layer, testing a number of neurons from 1 to 10. It was observed that six neurons give the lowest error on CV set.

5.6.4 Results and Discussions

Figure 13 shows the scatter plot of simulated discharge in verification period. For many data points HBV is quite accurate but its error (uncertainty) is quite high during the peak flows. This can be explained by the fact that the version of the HBV model used in this study is the lumped model and one cannot expect high accuracy from it.

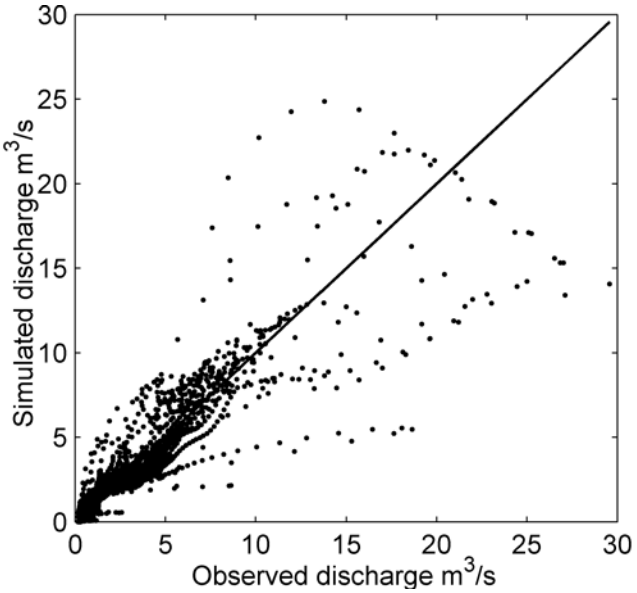


Figure 17: Scatter plot of observed and simulated discharge from HBV model

ANN-based uncertainty model *V* trained on the data generated by MCS, was tested on the verification data set; its performance is shown in Figure 14. Figure 15 shows the hydrograph with the 90% uncertainty bounds predicted by ANN together with the MC simulation uncertainty bounds in the verification period. It can be said that ANN reproduces the MC simulations uncertainty bounds reasonably well, in spite of the low correlation of the input variables with the PIs. Although some errors can be noticed, the predicted uncertainty bounds follow the general trend of MC uncertainty bounds. Noticeably the model fails to capture the observed flow during one of the peak events (bottom left figure). Note however, that the results of *V* and MC are visually closer to each other than both of them to the observed data.

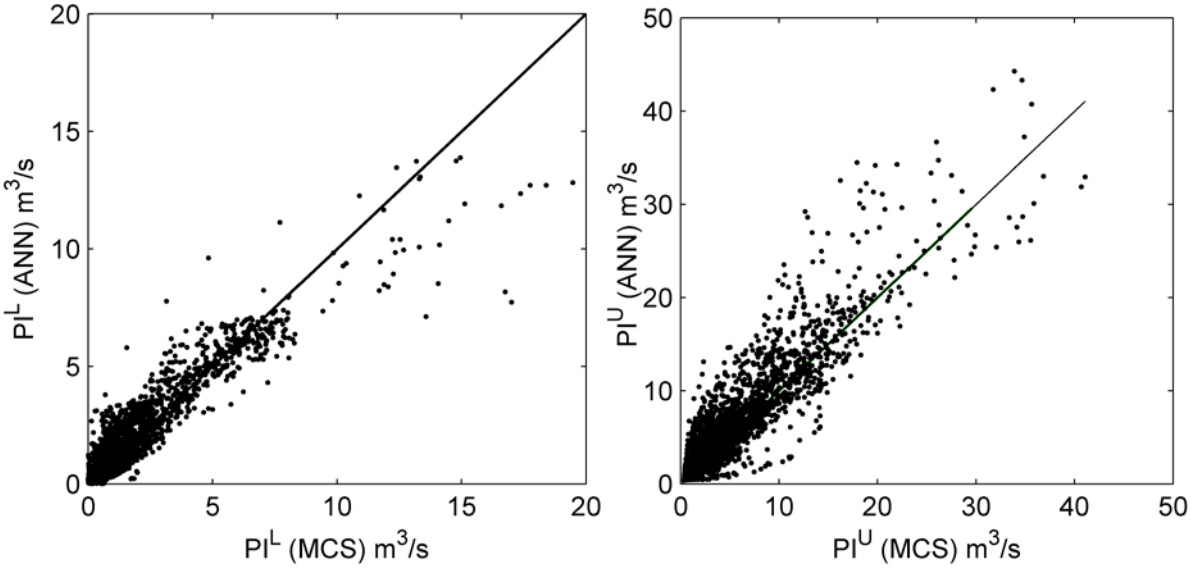


Figure 18: Scatter plots showing the performance of the ANN-based uncertainty model trained on the data generated by MCS, and estimating the prediction interval for the verification data set

Detailed analysis reveals that estimated uncertainty bounds contain 77.00% (*PICP*) of the observed runoffs, which is very close to the MC simulation result (77.24 %). The average width of prediction intervals (*MPI*) estimated by ANN is narrower (1.93 m^3/s) as compared

to the value obtained with MCS (2.09 m^3/s). Further analysis of the results reveals that 14.74% of the observed data are below the lower uncertainty bounds whereas 8.01% of data are above the upper bounds.

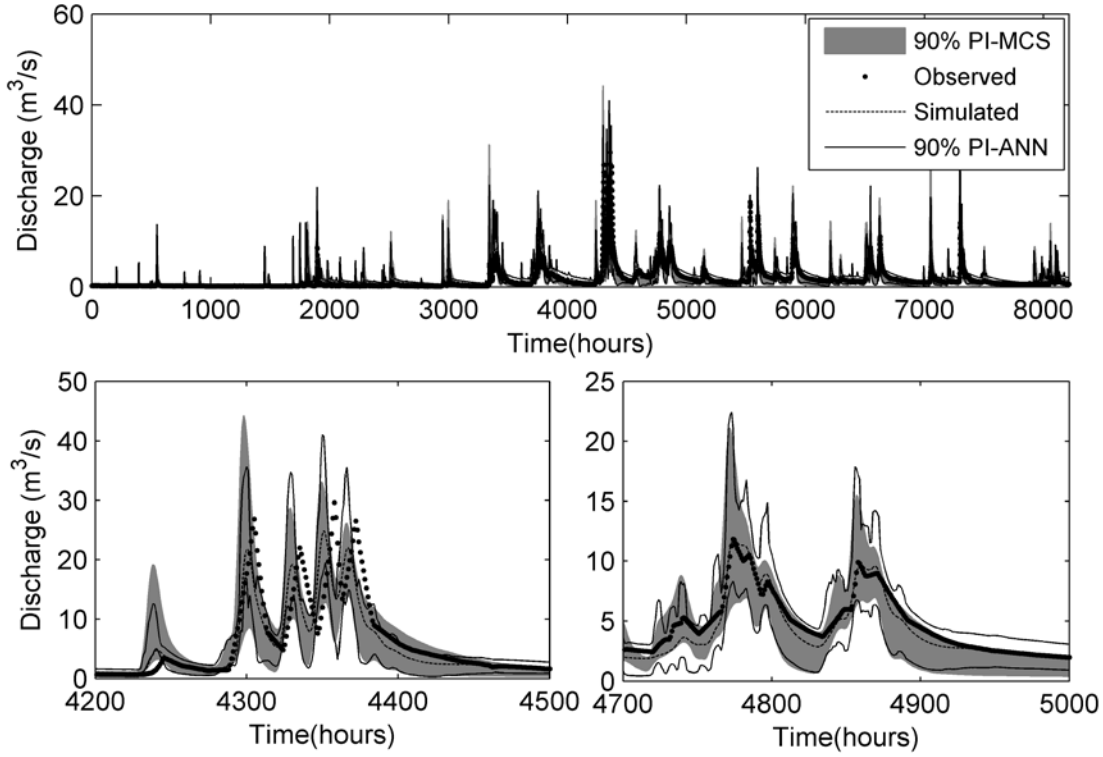


Figure 19: 90% prediction bounds estimated by MC simulation and ANN in verification period (1995/06/24 05:00 to 1996/05/31 13:00).

The predictive capability of ANN model in estimating lower and upper PIs are compared in both calibration and verification period and it appears that the correlation coefficient and RMSE for PI^L is higher than those of PI^U. This can be explained by the fact that PI^U corresponds to the higher values of flow (where the HBV model is less accurate) and has higher variability, which makes its prediction a difficult task.

In this example, the use of the MLUE method to replicate the results of Monte-Carlo simulations has been demonstrated. The method is computationally efficient and can be used in real time application when the large number of model runs required, and is applicable to various kinds of models.

6 Current trends in uncertainty studies

There are a number avenues along which research in model uncertainty analysis develops – they relate to the efficient (economical) but still informative methods of sampling, building better *predictive* models of uncertainty, attention to all sources of uncertianty and not only to parametric one, communication of uncertainty to decision makers, and, technologically, application of parallel computing. Two of these avenues are charactersied below.

6.1 Analysis of various sources of uncertainty

Most of the probabilistic techniques for uncertainty analysis treat only one source of uncertainty (i.e. parameter uncertainty). Recently, attention has been given to other sources of uncertainty, such as input uncertainty or structure uncertainty, as well as integrated approach to combine different sources of uncertainty. The research shows that input or structure uncertainty is more dominant than the parameter uncertainty. For example, Kavetski et al. (2006) and Vrugt et al. (2008a) among others treat input uncertainty in hydrological modelling using Bayesian approach. Butts et al. (2004) analysed impact of the model structure on hydrological modelling uncertainty for streamflow simulation. Recently new schemes have emerged to estimate the combined uncertainties in rainfall-runoff predictions associated with input, parameter and structure uncertainty. For instance, Ajami et al. (2007) used an integrated Bayesian multimodel approach to combine input, parameter and model structure uncertainty. Liu and Gupta (2007) suggested an integrated data assimilation approach to treat all sources of uncertainty. Spatial uncertainty is studied more and more as well; for example in flood inundation studies uncertainties in GIS (Brown and Heuvelink 2007) and in DEM (Moya et al., 2010) are to be taken into account.

Regarding the sources of uncertainty, Monte Carlo type methods are widely used for parameter uncertainty, Bayesian methods and/or data assimilation can be used for input uncertainty and Bayesian model averaging method is suitable for structure uncertainty. The appropriate uncertainty analysis method also depends on whether the uncertainty is represented as randomness or fuzziness. Similarly, uncertainty analysis methods for real time forecasting purposes would be different from those used for design purposes (e.g. when estimating design discharge hydraulic structure design).

6.2 Communication of uncertainty to decision makers and stakeholders

It should be noted that, the practice of uncertainty analysis and the use of the results of such analysis in decision making is not yet widely spread. Some possible misconceptions are stated by (Pappenberger and Beven, 2006):

> "*a) uncertainty analysis is not necessary given physically realistic models, b) uncertainty analysis is not useful in understanding hydrological and hydraulic processes, c) uncertainty (probability) distributions cannot be understood by policy makers and the public, d) uncertainty analysis cannot be incorporated into the decision-making process, e) uncertainty analysis is too subjective, f) uncertainty analysis is too difficult to perform and g) uncertainty does not really matter in making the final decision*".

Some of these misconceptions have however explainable reasons, so the fact remains: more has to be done in bringing the reasonably well developed apparatus of uncertainty analysis and prediction to decision making practice. In other words, *communication of uncertainty* to decision makers and other stakeholder is still a challenge.

7 References

Abebe, A.J., Guinot, V. and Solomatine, D.P. (2000). Fuzzy alpha-cut vs. Monte Carlo techniques in assessing uncertainty in model parameters. Proc. 4th Int. Conference on Hydroinformatics, Cedar Rapids.

Ajami, N.K., Duan, Q. and Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Research, 43, W01403, doi: 10.1029/2005WR004745.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. (1998). Crop evapotranspiration - guidelines for computing crop water requirements. Irrigation and Drainage Paper No. 56, FAO, Rome. Available at: http://www.fao.org/docrep/X0490E/x0490e00.htm.

Bárdossy, A. and Duckstein, L. (1995). Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological and Engineering Systems. CRC press Inc, Boca Raton, Florida, USA.

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6, pp. 279-298.

Beven KJ. (2001). Uniqueness of place and process representations in hydrological modeling. Hydrol. Earth Sys. Sci. 4, 203–213.

Braun, L.N. and Renner, C.B. (1992). Application of a conceptual runoff model in different physiographic regions of Switzerland. Hydrological Sciences Journal, 37(3), 217-232.

Brown, J.D., Heuvelink G.B.M., 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers & Geoscience* **33**, 172-190.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology, 298, 222-241.

Draper N R & Smith H. Applied regression analysis. New York: Wiley, 1966. 407 p.

Freedman, D.A. (2005) *Statistical Models: Theory and Practice*, Cambridge University Press. ISBN 978-0-521-67105-7.

Freer, J., Beven, K. and Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research, 32(7), pp. 2161-2173.

Y. Gan, Q. Duan, W. Gong, C. Tong, Y. Sun, W. Chu, A. Ye, C. Miao, Z. Di. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, 51 (2014) 269-285.

Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural networks and the bias/variance dilemma. Neural Computation, 4(1), 1-58.

Gouldby, B. and Samuels, P. (2005). Language of risk . Project definitions. FLOODsite Consortium Report T32-04-01. Available at: www.floodsite.net.

Gupta, H. V., Beven, K. J. and Wagener, T. (2005). Model calibration and uncertainty estimation. In: M. Andersen (ed.), Encyclopedia of Hydrological Sciences, John Wiley & Sons, New York, NY, USA, pp. 2015-2031.

Gupta H. V., Wagener T and Liu Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. Hydrol. Proc. 22(18), 3802–3813, doi:10.1002/hyp.6989.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1), 97-109.

Jacquin, A. P., and A. Y. Shamseldin (2007), Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, Water Resour. Res., 43, W04425, doi:10.1029/2006WR005072

Jacquin A. P. (2010). Possibilistic uncertainty analysis of a conceptual model of snowmelt runoff. *Hydrol. Earth Syst. Sci*., 14, 1681–1695.

Kavetski, D., Kuczera, G. and Franks, S.W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42, W03407, doi: 10.1029/2005WR004368.

Klir, G. and Folger, T. (1988). Fuzzy sets, uncertainty, and information. Prentice Hall, Englewood Cliffs, NJ, USA.

Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33-50.

Koenker, R.: Quantile Regression, Cambridge University Press, 2005.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J. of Hydrology, 211, 69-85.

Kundzewicz, Z. (1995). Hydrological uncertainty in perspective. In: Z. Kundzewicz (ed.), New Uncertainty Concepts in Hydrology and Water Resources, University Press, Cambridge, UK, pp. 3-10.

Larsen, Richard J. and Marx, Morris L. (2012). *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall.

LeBaron, B. and Weigend, A.S. (1994). Evaluating Neural Network Predictors by Bootstrapping, Proc. of Int. Conf. on Neural Information Processing (ICONIP'94), Seoul, Korea.

Letcher, R. A., Jakeman, A. J., and Croke, B. F. W. (2004). "Model development for integrated assessment of water allocation options." Water Resources Research, 40, W05502.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergtröm, S. (1997). Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology, 201, pp. 272-228.

Liu, Y. and Gupta, H.V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, 43, W07401, doi: 10.1029/2006WR005756.

D. Li, C. Liu, W. Gan (2099). A New Cognitive Model: Cloud Model. *Int. J. Intelligent Systems*, 24, 357–375.

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

Maskey, S.; Guinot, V.; Price, R. K. (2004). Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach, Advances in Water Resources, 27(9), 889-898.

McKay, M.D., Beckman, R.J. and Conover, W.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 239-245.

McLaughlin, M.P. (1999). The very game: a tutorial on mathematical modelling. www.geocities.com/~mikemclaughlin.

Melching, C.S. (1995). Reliability estimation. In: V.P. Singh (ed.), Computer Models of Watershed Hydrology, Water Resources Publications, Highlands Ranch, CO, USA, pp. 69-118.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1953). Equations of State Calculations by Fast Computing Machines. The Journal of Chemical Physics, 26(6), 1087-1092.

Montanari, A. (2007). What do we mean by uncertainty? The need for a consistent wording about uncertainty assessment in hydrology. Hydrological Processes, 21(6), 841-845.
Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resources Research, 40, W01106, doi: 10.1029/2003WR002540.
Moradkhani, H. and Sorooshian, S. (2008). General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In: S. Sorooshian et al. (eds.), Hydrological Modelling and the Water Cycle, Springer, Heidelberg, Germany, pp. 1-24.
Moya Quiroga, V., Popescu, I., Solomatine D.P., Bociort, L. (2011). Cloud and cluster computing in uncertainty analysis of integrated flood models. *J. Hydroinformatics* (submitted).
Neuman, S.P. 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stochastic Environmental Research and Risk Assessment, 17(5), 291-305.
D. Raje, P.P. Mujumdar (2010). Hydrologic drought prediction under climate change: Uncertainty modeling with Dempster–Shafer and Bayesian approaches Advances in Water Resources 33, 1176–1186.
Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29, 1586–1597.
Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. (2008). Global Sensitivity Analysis. The Primer, Wiley.
Schulz, K., Huwe, B., 1997. Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach. J. Hydrol. 201, 211 – 229.
Schulz, K., Huwe, B., 1999. Uncertainty and sensitivity analysis of water transport modelling in a layered soil profile using fuzzy set theory. J. Hydroinformatics, 1(2), 127-138.
Shafer, G. (1976). Mathematical theory of evidence Princeton University Press, Princeton, NJ, USA.
Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379-423, 623-656.
Shrestha, D.L. and Solomatine, D. (2008). Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. Intl. J. River Basin Manag., 6(2), 109-122.
Shrestha, D.L., Kayastha, N., and Solomatine, D. P.. A novel approach to parameter uncertainty analysis of hydrological models using neural networks. Hydrol. Earth Syst. Sci., 13, 1235–1248, 2009.
Solomatine D.P. (1999). Two strategies of adaptive cluster covering with descent and their comparison to other algorithms. *J. of Global Optimization*, 14(1), 55-78.
Solomatine D.P., Shrestha D.L. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Res*. 45, W00B11, doi:10.1029/2008WR006839, 2009.
Song X., Zhang J., Zhan C., Xuan Y., Ye M., Xu C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *J. Hydrology*, 523 (2015) 739–757.
Spear, R.C. and Hornberger, G.M. (1980). Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research, 14(1), 43-49.
Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, Water Resour. Res., 44, W00B06, doi:10.1029/2008WR006822.

Thiemann, M., Trosser, M., Gupta, H. and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. Water Resources Research, 37(10), 2521-2536.

Tung, Y.-K. (1996). Uncertainty and reliability analysis. In: L.W. Mays (ed.), Water Resources Handbook, McGraw-Hill, N.Y., USA, pp. 7.1-7.65.

Vrugt, J.A., Gupta, H.V., Bouten, W. and Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8), 1201, doi: 10.1029/2002WR001642.

Vrugt, J., ter Braak, C., Clark, M., Hyman, J. and Robinson, B. (2008a). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44, W00B09, doi: 10.1029/2007WR006720.

Vrugt J. A., ter Braak C. J. F., Gupta H. V., and Robinson B. A. (2008b). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment, 10.1007/s00477-008-0274-y.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M. and Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. of Nonlinear Sciences & Numerical Simulation* 10, 273-279.

Wagener, T., Boyle, D.P., Lees, M.J., Howard S. Wheater, H.S., Gupta, H.V. and Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1), 13–26.

Zimmermann, H.-J. (1997). A fresh perspective on uncertainty modeling: Uncertainty vs. uncertainty modeling. In: B.M. Ayyub and M.M. Gupta (eds.), Uncertainty analysis in engineering and sciences: Fuzzy logic, statistics, and neural network approach, Springer, Heidelberg, Germany, pp. 353-364.

Zadeh, L.A. (1965). Fuzzy sets. Information and Control, 8(3), pp. 338-353.

Zadeh, L.A. (1978). Fuzzy set as a basis for a theory of possibility. Fuzzy Sets and Systems, 1(1), pp. 3–28.