Uncertainty estimation in hydrology
Incorporating physical knowledge in stochastic modeling of uncertain systems

Alberto Montanari
Faculty of Engineering
University of Bologna
alberto.montanari@unibo.it

Demetris Koutsoyiannis
National Technical University of Athens
dk@ntua.itia.gr

Work carried out under the framework of the Research Project DATAERROR
(*Uncertainty estimation for precipitation and river discharge data.*
Effects on water resources planning and flood risk management)
Ministry of Education, University and Research - Italy

Premise: the problem of uncertainty estimation is a very old one….

- "*It seems to me that the condition of confidence or otherwise forms a very important part of the prediction, and ought to find expression*".

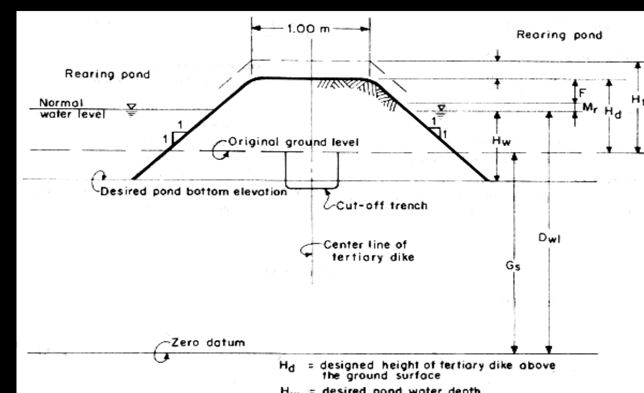
  W.E. Cooke, weather forecaster in Australia, 1905

- Hydraulic Engineers (fathers of hydrology) have been always well aware of uncertainty.

- Allowance for freeboards (safety factors) were always used to account for uncertainty in hydraulic engineering design.

- Expert judgement has been the main basis for hydrological uncertainty assessment in the past and will remain an essential ingredient in the future.

- Uncertainty in hydrology will never be eliminated (Koutsoyiannis et al., 2009).



Uncertainty in hydrology today: a fashion?

Google search for:

1) “uncertainty”:	→	32.400.000
2) “hydrology”:	→	34.800.000
3) “uncertainty” + “hydrology”:	→	2.210.000 6.4% of “hydrology 6.8% of “uncertainty”

ISI Web of Knowledge search in paper titles:

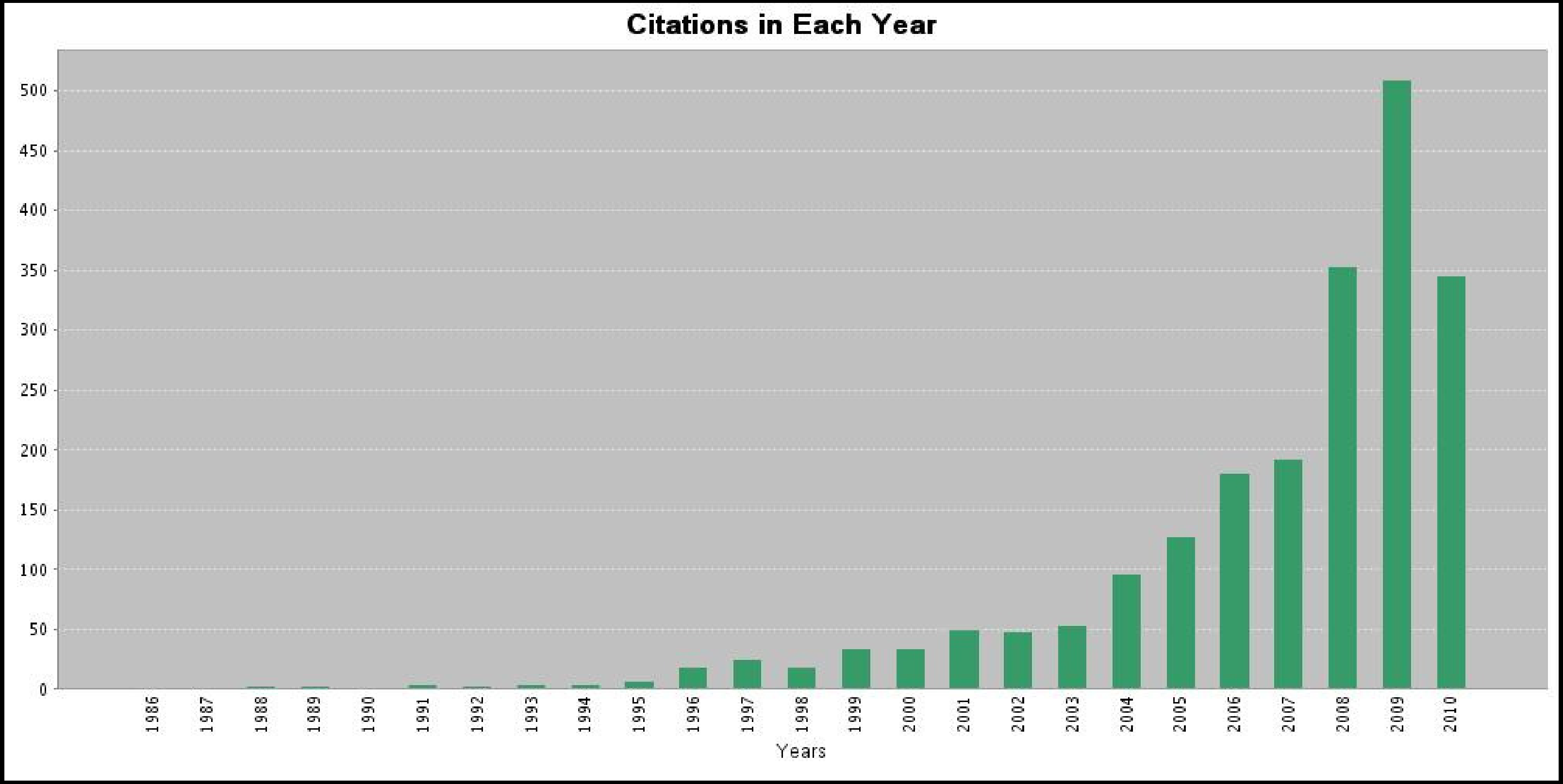
1)“hydrol*”:	→	46.123
2)“uncertainty” and “hydrol*”:	→	139

Most cited papers:

1)Beven K., Prophecy, reality and uncertainty in distributed hydrological modeling, Advances in water resources, 16, 41-51, 1993 (353 citations)

2)Vrugt J.A., Gupta H.V., Bouten W., Sorooshian S., A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resources Research, 39, 1201, 2003 (167 citations)

ISI Web of Knowledge search in paper titles for “uncertainty” and “hydrol*”



Deliverables obtained by research activity on uncertainty in hydrology

- **The working group** on uncertainty of the International Association of Hydrological sciences considered 25 methods for uncertainty assessment in hydrology (http://www.es.lancs.ac.uk/hfdg/uncertainty_workshop/uncert_methods.htm)

- **Matott et al**. (Water Resources Research, 2009) report 52 methods.

- **Many commentaries**: uncertainty assessment triggered several discussions. A very interesting debate was triggered by Beven (2006) in HP Today.

- **Key issue:** is statistical theory the appropriate tool to estimate uncertainty?

Drawbacks

- **Research** activity poorly structured.
- **Lack of clarity** about the research questions and related response.
- **Need for** a comprehensive theory

What is the practical question?

- Estimate uncertainty of hydrological simulation/prediction.
- Users typically want a confidence interval for simulation/prediction
- Confidence interval: for an assigned confidence level α, the confidence interval is the range around the model output encompassing the true data with a probability equal to α
- Probability: frequentist interpretation. An experiment is defined which is repeatable and the probability of an event is defined as the frequency of the event for number of trials tending to infinity.
- Probability: Bayesian interpretation. Probability is defined as the degree of belief about an event happening. It is subjective.
- If the experiment is well defined and Bayesian probability reliably estimated, the two definitions should give the same result.

- The above considerations clearly show that different methods for estimating confidence bands exist, even if we use just probability theory to compute them.

- Is it feasible to set up a theory?

What are the basic elements of a theory?

- **In science**, the term "theory" is reserved for explanations of phenomena which meet basic requirements about the kinds of empirical observations made, the methods of classification used, and the consistency of the theory in its application among members of the class to which it pertains. A theory should be the simplest possible tool that can be used to effectively address the given class of phenomena.

- **Basic elements of a theory**:
    - Subject.
    - Definitions.
    - Axioms or postulates (assumptions).
    - Basic principles.
    - Theorems.
    - Models.
    - Ethic principles.
    - …..

  - **Important:** a theory of a given subject is not necessarily unique

Towards a theory of uncertainty assessment in hydrology

- **Main subject**: estimating the uncertainty of the output from a hydrological model (global uncertainty).

- **Side subjects**: estimating data uncertainty (rainfall, river flows etc.), parameter uncertainty, model structural uncertainty, calibration, validation…. and more.

- **Two basic assumptions**:

1. **We assume** that global uncertainty is estimated through statistics and probability.
   This is not the only possible way to estimate uncertainty. Zadeh (2005) proposed to introduce a Generalized Theory of Uncertainty (GTU) encompassing all the possible methods to assess uncertainty, including probability theory and fuzzy set theory. Fuzzy set theory, in particular possibility theory, is an interesting opportunity for hydrology.
2. **We assume** that global uncertainty is formed up by:
   - Data uncertainty
   - Model parameter uncertainty
   - Model structural uncertainty

Why statistics?

- From my point of view, statistics is a set of tools to objectively profit from experience.

- Statistics, like any other method, is based on assumptions. Therefore there remains a certain degree of subjectivity.

- Statistical assumptions can largely be tested and must be tested.

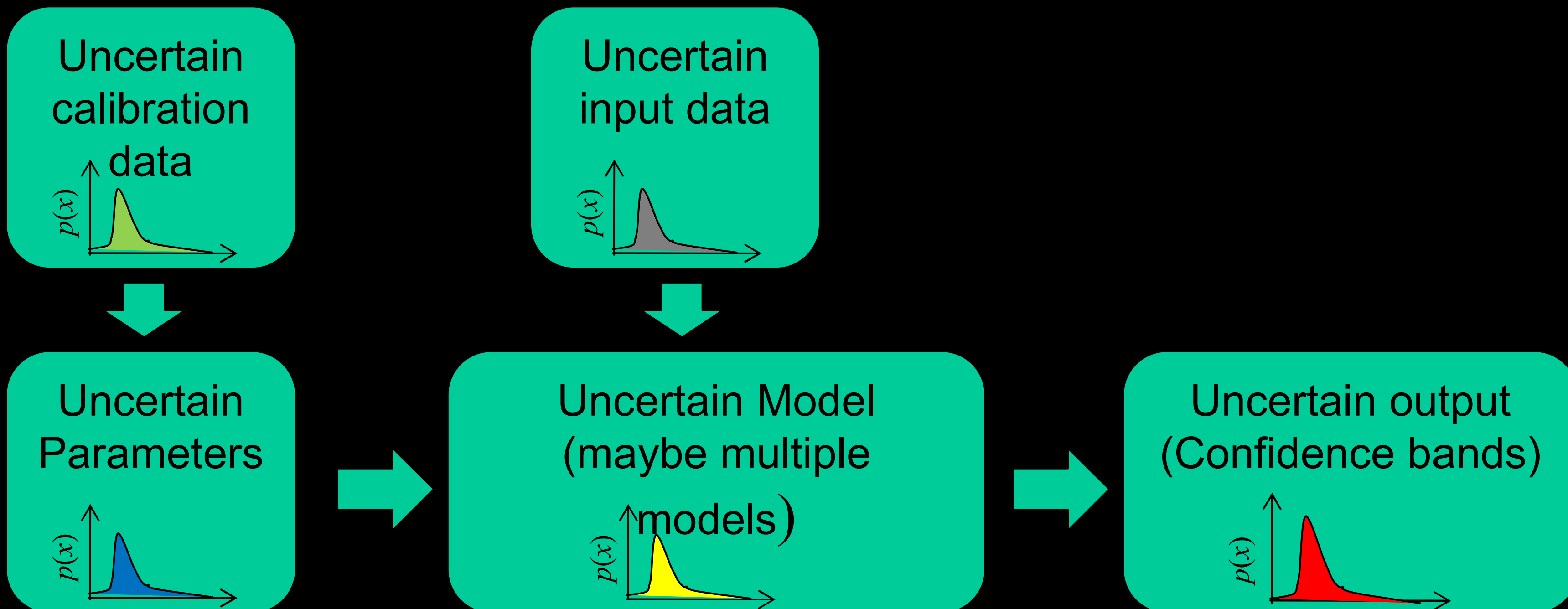
- Statistics can be used to model physically-based systems (see later) and has never been a synonym of lack of understanding.

- Statistics is often based on the assumption of stationarity which is questioned today. This debate is nonsensical. It is wrong to question the use of statistics by saying that systems are changing.

- Statistics has always been used to model the dynamics of changing and evolving systems (financial markets, internet traffic, etc).

- If a system is non-stationary, this unavoidably implies that non-stationarity can be deterministically described and therefore it can be embedded in statistical non-stationary model.

Towards a theory of uncertainty assessment in hydrology



Setting up hydrological models in a stochastic framework
A premise on terminology

Physically-based, spatially-distributed and deterministic are often used as synonyms. This is not correct.


Sir Isaac Newton
(1689, by Godfrey Kneller)

- Physically-based model: based on the application of the laws of physics. In hydrology, the most used physical laws are the Newton's law of the gravitation and the laws of conservation of mass, energy and momentum.
- Spatially-distributed model: model's equations are applied at local instead of catchment scale. Spatial discretization is obtained by subdividing the catchment in subunits (subcatchments, regular grids, etc).
- Deterministic model: model in which 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

A premise on terminology

Fluid mechanics obeys the laws of physics. However:

- Most flows are turbulent and thus can be described only probabilistically (note that the stress tensor in turbulent flows involves covariances of velocities).
- Even viscous flows are au fond described in statistical thermodynamical terms macroscopically lumping interactions at the molecular level.

It follows that:

- A physically-based model is not necessarily deterministic.

  A hydrological model should, in addition to be physically-based, also consider chemistry, ecology, etc.

  In view of the extreme complexity, diversity and heterogeneity of meteorological and hydrological processes (rainfall, soil properties…) physically-based equations are typically applied at local (small spatial) scale. It follows that:

- A physically-based model often requires a spatially-distributed representation.

A premise on terminology

In fact, some uncertainty is always present in hydrological modeling. Such uncertainty is not related to limited knowledge (epistemic uncertainty) but is rather unavoidable.

It follows that a deterministic representation is not possible in catchment hydrology.

The most comprehensive way of dealing with uncertainty is statistics, through the theory of probability.

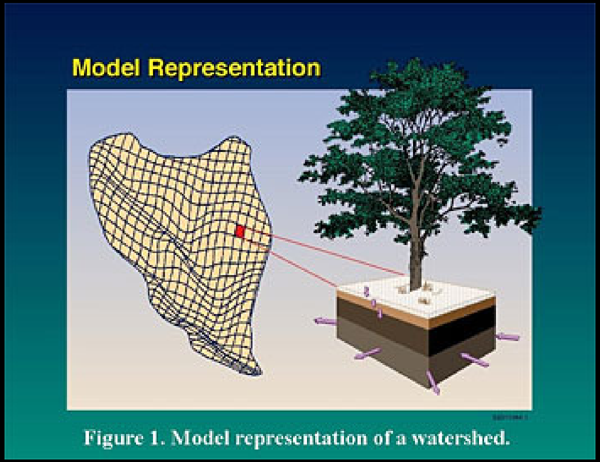


Figure 1. Model representation of a watershed.

Figure taken from http://hydrology.pnl.gov/

Therefore a stochastic representation is unavoidable in catchment hydrology (sorry for that...).

The way forward is the stochastic physically-based model, a classical concept that needs to be brought in new light.

A premise on stochastic models

"Stochastic" is a term that is very often used to mean the lack of a causal relationship between input and output, by often implying "lack of understanding".

~~Deterministic model~~ → ~~**Impossible for lack of understanding?**~~ → ~~Stochastic model~~

Understanding the dynamics of the process → **Is the process affected by uncertainty?** → Stochastic model (Expressing the dynamics in terms of probability)

Formulating a physically-based model within a stochastic framework

Hydrological model:

in a deterministic framework, the hydrological model is usually defined as a single-valued transformation expressed by the general relationship:

$$\boxed{Q_p = S\,(\varepsilon, \mathbf{I})}$$

where Q_p is the model prediction, S expresses the model structure, $\mathbf{I}$ is the input data vector and ε the parameter vector.

In the stochastic framework, the hydrological model is expressed in stochastic terms, namely (Koutsoyiannis, 2010):

$$\boxed{f_{Qp}\,(Q_p) = K\, f_{\varepsilon, I}(\varepsilon, \mathbf{I})}$$

where f indicates the probability density function, and K is a transfer operator that depends on model S.

Formulating a physically-based model within a stochastic framework

Assuming a single-valued (i.e. deterministic) transformation $S(\varepsilon, \mathbf{I})$ as in previous slide, the operator K will be the Frobenius-Perron operator (e.g. Koutsoyiannis, 2010).

However, K can be generalized to represent a so-called stochastic operator, which corresponds to one-to-many transformations S.

A stochastic operator can be defined using a stochastic kernel $k(e, \boldsymbol{\varepsilon}, \mathbf{I})$ (with e intuitively reflecting a deviation from a single-valued transformation; in our case it indicates the model error) having the properties

$$k(e, \boldsymbol{\varepsilon}, \mathbf{I}) \geq 0 \quad \text{and} \quad \int_e k(e, \boldsymbol{\varepsilon}, \mathbf{I})\, de = 1$$

Formulating a physically-based model within a stochastic framework

Specifically, the operator K applying on $f_{\boldsymbol{\varepsilon},\,\mathbf{I}}\,(\boldsymbol{\varepsilon}, \mathbf{I})$ is then defined as (Lasota and Mackey, 1985, p. 101):

$$K\, f_{\boldsymbol{\varepsilon},\,\mathbf{I}}(\boldsymbol{\varepsilon}, \mathbf{I}) = \int_{\boldsymbol{\varepsilon}} \int_{\mathbf{I}} k(e, \boldsymbol{\varepsilon}, \mathbf{I})\, f_{\boldsymbol{\varepsilon},\,\mathbf{I}}\,(\boldsymbol{\varepsilon}, \mathbf{I})\, d\boldsymbol{\varepsilon}\, d\mathbf{I}$$

If the random variables $\boldsymbol{\varepsilon}$ and $\mathbf{I}$ are independent, the model can be written in the form:

$$f_{Q_p}(Q_{\mathrm{p}}) = K\,[f_{\boldsymbol{\varepsilon}}(e)\, f_{\mathbf{I}}\,(\mathbf{I})]$$

$$f_{Q_p}\,(Q_{\mathrm{p}}) = \int_{\boldsymbol{\varepsilon}} \int_{\mathbf{I}} k(\mathrm{e}, \boldsymbol{\varepsilon}, \mathbf{I})\, f_{\boldsymbol{\varepsilon}}\,(\varepsilon)\, f_{\mathbf{I}}\,(\mathbf{I})\, d\boldsymbol{\varepsilon}\, d\mathbf{I}$$

Formulating a physically-based model within a stochastic framework

Estimation of prediction uncertainty:

Further assumptions:

1) model error is assumed to be independent of input data error and model parameters.
2) Prediction is decomposed in two additive terms, i.e. :

$$Q_p = S(\boldsymbol{\varepsilon}, \mathbf{I}) + e$$

where S represents the deterministic part and the structural error e has density $f_e(e)$.

3) Kernel independent of $\boldsymbol{\varepsilon}$, $\mathbf{I}$ (depending on e only), i.e.:

$$k(e, \boldsymbol{\varepsilon}, \mathbf{I}) = f_e(e)$$

By substituting in the equation derived in the previous slide we obtain:

$$\boxed{f_{Q_p}(Q_p) = \int_{\boldsymbol{\varepsilon}} \int_{\mathbf{I}} f_e(Q_p - S(\boldsymbol{\varepsilon}, \mathbf{I}))\, f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})\, f_{\mathrm{I}}(\mathbf{I})\, d\boldsymbol{\varepsilon}\, d\mathbf{I}}$$

Formulating a physically-based model within a stochastic framework

Symbols:

- Q_p — Predicted value of the true hydrological variable
- $S(\varepsilon,\mathbf{I})$ — Deterministic hydrological model
- e — Model structural error
- ε — Model parameter vector
- $\mathbf{I}$ — Input data vector

From the deterministic formulation:

$$\boxed{Q_p = S(\boldsymbol{\varepsilon}, \mathbf{I})}$$

to the stochastic simulation:

$$\boxed{f_{Q_p}(Q_p) = \int_{\boldsymbol{\varepsilon}} \int_{\mathbf{I}} f_e(Q_p - S(\boldsymbol{\varepsilon}, \mathbf{I}))\, f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})\, f_{\mathbf{I}}(\mathbf{I})\, d\boldsymbol{\varepsilon}\, d\mathbf{I}}$$

Formulating a physically-based model within a stochastic framework

An example of application: model is generic and possibly physically-based. Let us assume that input data uncertainty can be neglected, and that probability distributions of model error and parameters are known.

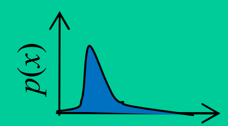
Repeat j times

Pick up a parameter vector ε from the model parameter space accordingly to probability $f_\varepsilon(\varepsilon)$

Input data vector (certain)

Compute model output and add n realisation of model error from probability distribution $f_e(e)$

Obtain $n \cdot j$ points lying on $f_{Qp}(Q_p)$ and infer the probability distribution



Problems:
1)computational demands;
2)estimate $f_\varepsilon(\varepsilon)$ and $f_e(e)$

Placing existing techniques into the theory's framework

- **Generalised Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992)**:

✓The most used method for uncertainty assessment in hydrology:
Google Scholar search for "Generalised likelihood uncertainty": 350 papers

✓It has often been defined as an "informal" statistical method

✓Criticised for being subjective and therefore not coherent (Christensen, 2004; Montanari, 2005; Mantovan and Todini, 2006; Mantovan et al., 2007)

✓Successfully applied by many researchers (Aronica et al., 2002; Borga et al., 2006; Freni et al., 2009)

Placing existing techniques into the theory's framework

Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology

Keith Beven*, Jim Freer

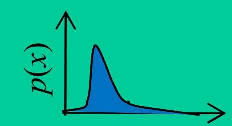
Beven and Freer, 2001

$f_{Qp}(Q_p)$ is computed by rescaling an informal likelihood measure for the model (usually a goodness of fit index)

Repeat j times

Compute model output Q_p, compute model likelihood L(ε) and obtain an estimate of $f_{Qp}(Q_p)$

Obtain j points lying on $f_{Qp}(Q_p)$ and infer the related probability distribution



Pick up a parameter vector from the model parameter space accordingly to probability $f(\varepsilon)$ (uniform distribution is often used)

Input data vector (certain)

Problems:
1) computational demands;
2) informal likelihood and rescaling method are subjective

Placing existing techniques into the theory's framework

- **Bayesian Forecasting systems (BFS; Krzysztofowicz, 2002)**:

✓Described in a series of papers by Krzysztofowicz and other published from 1999 to 2004.

✓It has been conceived to estimate the uncertainty of a river stage (or river flow) forecast derived through a rainfall forecast and a hydrological model as a mean to transform precipitation into river stage (or river flow).

✓Basic assumption: dominant source of uncertainty is rainfall prediction. Parameter uncertainty and data uncertainty implicitly accounted for.

Placing existing techniques into the theory's framework

Bayesian Forecasting System (BFS)

Bayesian system for probabilistic river stage forecasting

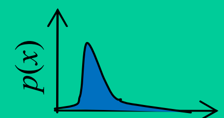
Roman Krzysztofowicz*

Krzysztofowicz, 2002

$f_{Qp}(Q_p \mid S(\varepsilon,\mathbf{I}))$ is computed by assuming that $f_{Qp}(Q_p, S(\varepsilon,\mathbf{I}))$ is bivariate meta-Gaussian

Compute model output Q_p, and compute $f_{Qp}(Q_p \mid S(\varepsilon,\mathbf{I}))$ from historical model runs

Obtain $f_{Qp}(Q_p)$



Parameter vector (certain)

Input data vector (certain)

Problems:
1)The bivariate meta-Gaussian distribution hardly provides a good fit

Placing existing techniques into the theory's framework

- **Meta-Gaussian approach (Montanari and Brath, 2004; Montanari and Grossi, 2008)**:

✓Data and parameter uncertainty implicitly accounted for.

✓It has been conceived to estimate the uncertainty of an optimal rainfall-runoff model with an optimal parameter set.

✓Basic assumption: joint distribution of model prediction and model error is bivariate meta-Gaussian.

Placing existing techniques into the theory's framework

Meta-Gaussian Approach

A stochastic approach for assessing the uncertainty of rainfall-runoff simulations

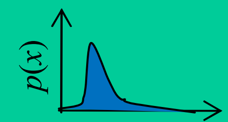
Alberto Montanari and Armando Brath
Faculty of Engineering, University of Bologna, Bologna, Italy

Montanari and Brath, 2004

$f_e\ (e|Q_p)$ is computed by assuming that $f_e(e,Q_p)$ is bivariate meta-Gaussian

Compute model output Q_p, and compute $f(e|\ Q_P)$ from historical model runs

Obtain $f_{Qp}(Q_p)$



Parameter vector
(certain)

Input data vector
(certain)

Problems:
1)Needs to be calibrated with long series of historical model runs

Example: Leo Creek at Fanano (Italy)

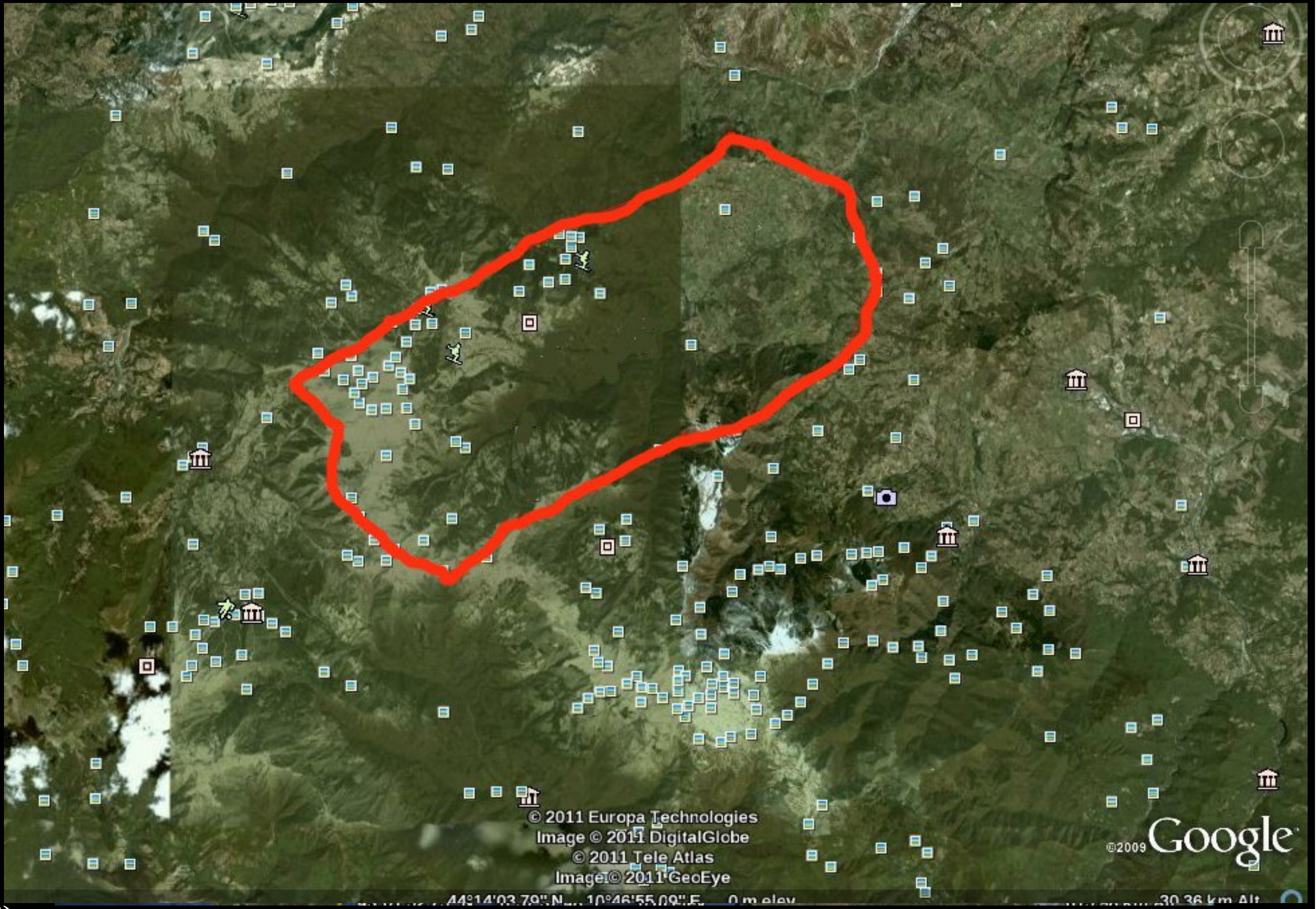
Basin area:
35 km^2

Main stream length:
10 km

Max altitude:
1768 m slm

Mean flow:
5 m^3/s

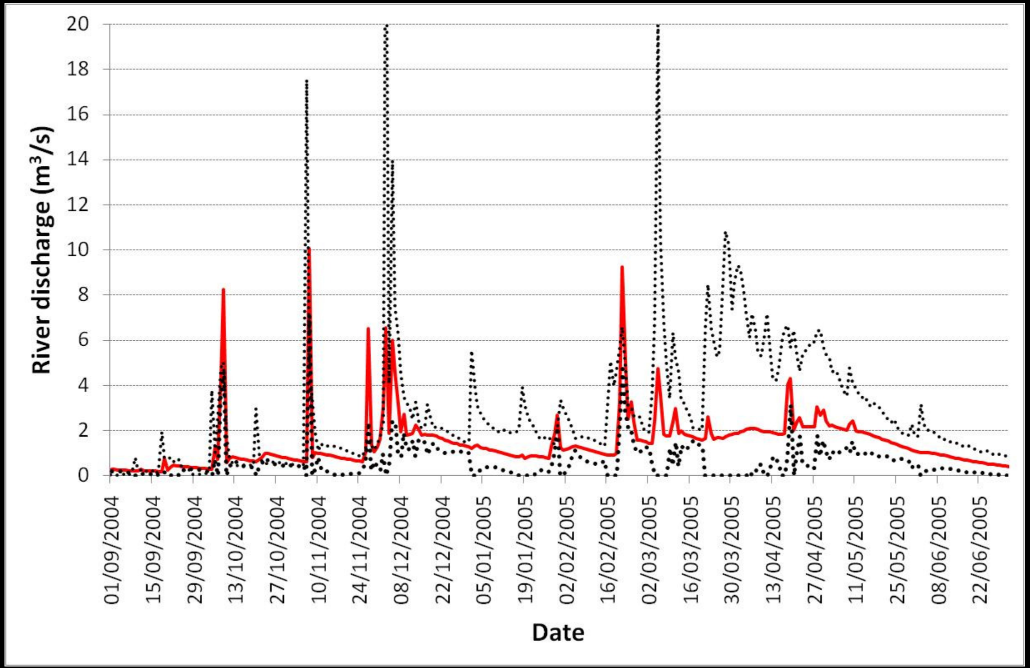
Calibration:
NE=0.62



(Courtesy by: Elena Montosi)

Estimation of the predictive distribution

Rainfall runoff model: AFFDEF – Daily time scale – Conceptual , 7 parameters
Parameter distribution: estimated by using DREAM (Vrugt and Robinson, 2007)
Generation of random samples of model error: by using the meta – Gaussian approach (Montanari and Brath, 2004; Montanari and Grossi, 2008).



Research challenges

To include a physically-based model within a stochastic framework is in principle easy. Nevertheless, relevant research challenges need to be addressed:

- numerical integration (e.g. by Monte Carlo method) is computationally intensive and may result prohibitive for spatially-distributed models. There is the need to develop efficient simulation schemes;
- a relevant issue is the estimation of model structural uncertainty, namely, the estimation of the probability distribution $f(e)$ of the model error. The literature has proposed a variety of different approaches, like the GLUE method (Beven and Binley, 1992), the meta-Gaussian model (Montanari and Brath, 2004; Montanari and Grossi, 2008), Bayesian Model Averaging. For forecasting, Krzysztofowicz (2002) proposed the BFS method;
- estimation of parameter uncertainty is a relevant challenge as well. A possibility is the DREAM algorithm (Vrugt and Robinson, 2007).

Concluding remarks

- A deterministic representation is not possible in hydrological modeling, because uncertainty will never be eliminated. Therefore, physically-based models need to be included within a stochastic framework.
- The complexity of the modeling scheme increases, but multiple integration can be easily approximated with numerical integration.
- The computational requirements may become very intensive for spatially-distributed models.
- How to efficiently assess model structural uncertainty is still a relevant research challenge, especially for ungauged basins.

- MANY THANKS to: Guenter Bloeschl, Keith Beven, Elena Montosi, Siva Sivapalan, Francesco Laio

http://www.albertomontanari.it - alberto.montanari@unibo.it

Main References (not all those cited in the text)

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

Koutsoyiannis, D., A random walk on water, Hydrology and Earth System Sciences Discussions, 6, 6611–6658, 2009.

Krzysztofowicz, R., Bayesian system for probabilistic river stage forecasting, Journal of Hydrology, 268, 16–40, 2002.

Lasota, D.A., Mackey, M.C., Probabilistic properties of deterministic systems, Cambridge Universityy Press, 1985.

Montanari, A., Brath, A., A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. Water Resources Research, 40, W01106, doi:10.1029/2003WR002540, 2004.

Montanari, A., Grossi, G., Estimating the uncertainty of hydrological forecasts: A statistical approach. Water Resources Research, 44, W00B08, doi:10.1029/2008WR006897, 2008.

Vrugt, J.A and Robinson, B.A., Improved evolutionary optimization from genetically adaptive multimethod search, Proceedings of the National Academy of Sciences of the United States of America, 104, 708-711, doi:10.1073/pnas.06104711045, 2007.

http://www.albertomontanari.it - alberto.montanari@unibo.it