Elsevier Editorial System(tm) for Advances in Water Resources
Manuscript Draft

Manuscript Number: AWR-11-360R1

Title: Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty

Article Type: SI: 35th anniversary

Keywords: hydrology, observation uncertainty, monitoring techniques, model complexity

Corresponding Author: Dr. Giuliano Di Baldassarre,

Corresponding Author's Institution: UNESCO-IHE

First Author: Alberto Montanari

Order of Authors: Alberto Montanari; Giuliano Di Baldassarre

Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty

Alberto Montanari[1], Giuliano Di Baldassarre[2*]

[1]Department DICAM, University of Bologna, Bologna, Italy (alberto.montanari@unibo.it)

[2]UNESCO-IHE Institute for Water Education, Delft, The Netherlands (g.dibaldassarre@unesco-ihe.org)

*Corresponding author: Giuliano Di Baldassarre, UNESCO-IHE Institute for Water Education, Department of Integrated Water Systems and Governance, Westvest 7, PO Box 3015, 2601 DA, Delft, The Netherlands, Phone: +31 (0)15 2151 846, g.dibaldassarre@unesco-ihe.org

Response to Reviewers' Comments

We thankfully acknowledge the three Anonymous Reviewers for providing very constructive and useful comments that enabled us to perform additional research work (see completely new Table 1 and Figure 1), improve the clarity and quality of the presentation of our work, and correct some inaccuracies that were present in the original version of our manuscript. The next three sections reports a detailed description of how we addressed all the comments and suggestions of each reviewer. The following notation was used:

Reviewer's comment

Reply from the authors

Response to Reviewer #1

General Comments:

This paper discusses the impact of data uncertainty on hydrological modeling and in particular on flood risk assessment studies. It is not quite clear what type of paper this is; considering its content, it seems to be an opinion paper; this is supported by the end of the introduction where it is stated that "the purpose of this paper is to illustrate the (..) effects of data uncertainty (..)" and to "indicate some potential ways forward to reduce this type of uncertainty". But the abstract as well as the conclusion are formulated as if it was a research paper. Given that there are only some simple examples for illustrative purposes, I would, however, not consider this contribution as a research paper. Furthermore, the content of the paper is confusing. Section 2 gives an overview of existing literature on river flow uncertainty and flood risk modeling (with some relevant literature missing, see detailed comments). Section 3 discusses to some extent the relation between model complexity and model sensitivity to river flow errors (with some simple examples). Section 4 gives some ideas about how to reduce the river flow uncertainty. None of the mentioned topics is developed in sufficient detail for a research paper. Overall, it appears to me that the paper contains interesting ideas (see also hereafter) but nothing is developed in sufficient detail for a research paper. It seems a bit like the paper had been written at a too early stage (see also p. 9, line 24: " We are currently testing several models with increasing complexity applied to different case studies.").

We agree with Reviewer#1. The structure of the original manuscript was not appropriate and some comments could sound either speculative or misleading. We also recognize that we did not make clear whether our manuscript was meant to be an opinion or a research article. This is why we strongly revised our manuscript and carried out additional numerical experiments (see completely new Table 1 and Figure 1) to better explore the impact of observation uncertainty on hydrological modelling in view of model complexity (see below).

The revised manuscript is to be considered a research paper.

Detailed comments:

- the paper seems to use the term "observation errors" as a synonym for "river flow errors"; I recommend reformulating at all instances since observation errors include e.g. errors on observed rainfall, temperature etc.

The Reviewer is right. The original version only looked at river flow errors. However, the revised version refers to observation uncertainty as we performed additional experiments were rainfall errors were also considered. In particular, the revised manuscript states:

"To investigate the impact of data errors on model performance, calibration/validation experiments were repeated by referring to corrupted river flow and rainfall data. Corruption was obtained by introducing 4 types of data errors, which are described here below.

- Corruption 1 – Random error in river flow data: it was obtained by multiplying each river flow observation by a random outcome from a uniform probability distribution in the range 0.8-1.2. The magnitude of the introduced error is large, but comparable with what was detected by previous analyses with respect to state of the art monitoring techniques [9,11,13].
- Corruption 2 – Systematic bias in flood flow data [14,15]: it was obtained by multiplying the river flows greater than 150 m^3/s by 1.25. This is a large but not unreasonable error that might be originated by extrapolation error in the rating curve (see above).
- Corruption 3 – Rainfall error: it was obtained by introducing a random error in the raingauge weights that are used to estimate mean areal rainfall. In detail, the weight of each raingauge was randomly picked up, at each time step, in the range ±20% of its optimal value according to a uniform probability distribution. The obtained weights were rescaled so that their cumulative sum is equal to one.
- Corruption 4 – Correlated noise in river flow data: it was obtained by multiplying each river flow observation by a correlated noise generated by using an autoregressive first-order linear stochastic process, with autoregressive coefficient equal to 0.5. The resulting noise has a unit mean and a standard deviation of 0.1, namely, statistics that are consistent with errors that are typically encountered in real world data [9,11,13]."

- start section 3: According to the introduction and the abstract, the paper focuses on the impact of river flow uncertainty on flood risk modeling arguing that this has rarely been done, which is certainly true; in exchange, I do not agree with the more general statement (start section 3) that "the impact of observation errors on hydrological modeling has been poorly investigated". As far as I see, there is a huge amount of literature on this (e.g. studies on the relation between uncertainty rainfall estimates and model uncertainty, e.g. the work of Bardossy et al.). The same holds for model identification in the face of data uncertainty (Vrugt et al, Beven et al., Freer et al., Kavetski et al); I recommend not mixing the two topics: effect of river flow uncertainty on flood risk estimation and the effect of data uncertainty (including input uncertainty!) on hydrological model identification (a far wider topic).

The Reviewer is right. The original manuscript included inappropriate statements and did not clarify the original points of our research work. The revised manuscript clarifies the goal of our study in view of the state of the art. In particular, the revised manuscript states:

"The above summary shows that considerable research work was devoted to estimate observation uncertainty in hydrology [9-17]. However, little attention was devoted to the most appropriate modelling approaches to be adopted to reduce its impact. For instance, the literature dedicated ample room to comparing the performance of different models (lumped versus distributed, conceptual versus physically-based, e.g. [18]), but the role of model complexity in the face of data uncertainty has remained largely unexplored. Experiences carried out in other scientific fields highlight that efficient solutions can be identified to minimise the effect of noise or bias in the data. For instance, in statistics a lot of attention is devoted to asymptotical properties of estimators to infer the probability distribution of parameters and results. Given that uncertainty is always present in hydrological modelling and will never be eliminated [19], we believe that there is the need for developing guidelines to drive the selection of the most appropriate model and calibration strategy depending on observation uncertainty and the actual information content of data.

The purpose of this paper is to provide a first contribution to identify best modelling practices in hydrology to minimise the impact of observation uncertainty. In detail, we argue that an appropriate selection of hydrological model complexity and calibration strategy can increase the robustness of hydrological studies against data errors. We test the above hypothesis by performing an extended set of calibration/validation experiments. To this end, we used synthetic

hydrological data that were corrupted by errors, emulating the uncertainty typically affecting hydrological records [9, 11, 13-15]."

- conclusion: Again, I don't think that it holds that the "impact (of data uncertainty) remains largely unexplored". The huge amount of hydrologic literature on modelling uncertainty contains many examples discussing the effect of data uncertainty. Here you focus on the uncertainty of river flow observations rather than on input uncertainty. This should be made clear.

The revised manuscript made clear that the main elements of novelty of this paper are the exploration of the impact of data uncertainty in view of model complexity and the indication of modelling strategies to reduce this impact. In particular, that part of the conclusions was revised as follows:

"Observation uncertainty is increasingly recognized to significantly affect hydrological modelling. A number of studies have attempted to estimate data errors and evaluate their impact on hydrological modelling. However, the actual impact of observation uncertainty in view of model complexity remains largely unexplored. Also, guidelines about the most appropriate modelling approaches to reduce the impact of the imprecision of hydrological data are still missing."

- conclusion: as far as I know, there is considerable literature on the estimation of model parameters given input / model / output uncertainty; thus: what is new about your conclusion that we need to "perform hydrological model identification and calibration in the face of observation errors"?)

As mentioned, the revised paper fully recognizes previous studies in the field with a large number of specific references. The novelty of our study is the investigation of the role of model complexity in reducing the impact of observation uncertainty. The introduction, discussion and conclusions were strongly revised accordingly (see above and revised manuscript).

- Conclusion: the idea of appropriate choice of model complexity to reduce impact of observational uncertainty is interesting and rather new (most papers on model complexity deal with the question of parameter identifiability but not with minimizing the impact of observational uncertainty); this paper would certainly be a valuable contribution if you could develop it around this idea

We acknowledge the Reviewer for this comment. Indeed, we followed her/his precious suggestion and carried out several numerical experiment to explore the selection of appropriate model complexity to reduce the impact of observational uncertainty (see above).

- Conclusion/abstract: is the sentence "the best modelling option to follow should be identified by (merely) relying on common sense, intuition and/or expert knowledge" not contradictory to the following sentence from the abstract?: "This paper shows how the effect of data uncertainty on hydrological modelling is difficult to predict by simply relying on intuition, common sense and expert knowledge"

The Reviewer is right. The first sentence was not properly formulated. This was corrected in the revised manuscript.

- literature: some relevant literature is missing, e.g. the work on total modelling error quantification by Kavetski et al., or the paper " Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, by McMillan et al., 2010, HP

We thank the Reviewer for suggesting interesting papers on the topic. We followed her/his advice and added the following three references (see revised manuscript):

"[40] Kavetski, D., G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005WR004368, 2006a.

[41] Kavetski, D., G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42, W03408, doi:10.1029/2005WR004376, 2006b.

[42] McMillan H, Jackson B, Clark M, Kavetski D, Woods R. Rainfall uncertainty in hydrologic modelling: An evaluation of multiplicative error models, Journal of Hydrology, 400, 83-94, 2011."

Response to Reviewer #2

Reviewer #2: The article has a somewhat unconventional structure in that it presents a narrative without worrying too much about scientific method. What I understand is that through a case study it becomes clear that measurement errors affect simple models more than more complex models. This is an interesting finding but I do not think that it is the first of its kind or that it is unique. The first author was editor of a special issue of WRR where much was explained about the role of uncertainty, also in measurement errors. There is a relation between measurement error, the information contained in the measurements (number of independent measurements), and the optimal complexity. The above mentioned special issue already contains an article concerning optimal model complexity, clearly related to measurement error. In computer science, this phenomenon is well studied and has led, for example, to the rather elegant theory of Vapnik. It is a bit difficult to define good advice concerning the article. For starters, I would recommend to re-write following a more orthodox structure: Intro, Methods&Materials, Results, Conclusions, Discussion. I know this may sound boring but it prevents the authors from just telling a speculative story. For example, page 11 lines 12-14: What does this mean? How can I use this/repeat this?

Indeed, the are many papers dealing with the topic, but this is now fully recognized in the revised manuscript, including a number of specific references.

For what concerns the WRR paper, (although there might be a misunderstanding as the Reviewer did not indicate a precise reference) that study is looking at hydrological measurements used as input data. Our study is completely different. We dealt with the uncertainty of the observations used to calibrate hydrological models and investigated the role of model structure in reducing the impact of such observation uncertainty.

Lastly, it is our belief that (to avoid, for instance, what Taleb calls the "ludic fallacy"), before extending elegant theories, a number of more modest and empirical studies are needed.

Anyhow, we fully recognize the limitation of our original manuscript (see above). We performed additional analyses (see new Table 1 and Figure 1) and the paper was strongly restructured. All the statements sounding like speculative stories were entirely removed (see revised manuscript).

At page 10, we have had our introduction (errors may be relevant) and the main results (errors affect a simple model more than a less simple model). For some reason, the conclusion is said to be non-intuitive but anyone who has gone into some depth concerning model complexity and data availability would have a different intuition...

We agree with Reviewer#2. Someone might have a different intuition. However, according to our experience, many expert modelers often (mis?)claim the principle of parsimony to support the use of simpler models when data errors are thought to be relevant. Their argument is that complex models may fit the noise in the data rather than the "real signal".

Anyhow, as mentioned, we do not aim at theorizing.

Our point is that the role of model complexity in view of observation uncertainty should be carefully examined (not relying on common sense, intuition or elegant theories).

To make this point clear, we revised that statement as follows (see revised manuscript):

"The above results indicate that the sensitivity of hydrological models to observation uncertainty should be carefully examined to identify the optimal model complexity in the face of observation uncertainty, because the actual results may contradict intuition and common sense."

At one point the authors suggest that observation errors reduce information content. True, but there are very good measures for this. Information is well defined within computer science through information theory, and so is the effect of measurement errors on information content.

We acknowledge again the Reviewer for suggesting elegant theories. However, as mentioned, our objective is to explore via simulation experiments the role of model complexity in limiting or amplifying the impact of observation errors.

After page 10 follows a narration concerning the advantages and disadvantages of different stage/discharge curve establishment methods. This part is really rather speculative and has no clear link to the content of the first part. In a way, this should be part of a different article.

We followed the Reviewers suggestion. As mentioned, the structure of the paper was strongly revised and this part, which shows possible strategies to reduce observation uncertainty, was significantly shortened.

I think that the main point of the article is that there is necessary (but not sufficient) proof that measurement error affects optimal model complexity. I would suggest to make this the main point and link that to established theory. I would be happy to look at any further iterations in this direction.

The main point of the article is to explore whether the choice of appropriate model structures can help reduce the impact of observation uncertainty. To this end, we carried out extended numerical exercises by referring to different types of data errors (see above).

Response to Reviewer #3

Reviewer #3: This paper deals with various issues on data uncertainty applied to hydrological modeling aka flood risk assessment studies. The paper has potential, but falls short of its aims as it is not either a review or a research paper as is. Some issues touched, however, are interesting and insightful. Rather than dusmissing it, as I was tempted to recommend, I would rather give the authors the chance to revise it -- thoroughly in the presentation and organization of material, especially clearly stating from the onset its scopes. I also have the impression that the literature search could be made more thorough. Moreover, the currently tested models applied to different case studies quoted in the text should be integral part of the revised manuscript. If the authors decide to do so, I would be willing to review it as a research paper.

We acknowledge the Reviewer for this constructive comment. Indeed, we followed her/his precious suggestion to make our manuscript a full research article (see above) and carried out several numerical experiment (see new Table 1 and Figure 1) to explore the selection of appropriate model complexity to reduce the impact of observational uncertainty.

*Highlights (for review)

Highlights of the paper:

"Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty"

Alberto Montanari & Giuliano Di Baldassarre

- We review the recent literature in the field of uncertainty in hydrological data
- We discuss how this observation uncertainty may undermine the reliability of hydrological predictions
- We show that the effect of data errors on hydrological model should be carefully evaluated in view of model complexity
- We show interesting results in the role of model structures in limiting or amplifying the impact of observation uncertainty

Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty

Alberto Montanari[1], Giuliano Di Baldassarre[2*]

[1]Department DICAM, University of Bologna, Bologna, Italy (alberto.montanari@unibo.it)

[2]UNESCO-IHE Institute for Water Education, Delft, The Netherlands (g.dibaldassarre@unesco-ihe.org)

*Corresponding author: Giuliano Di Baldassarre, UNESCO-IHE Institute for Water Education, Department of Integrated Water Systems and Governance, Westvest 7, PO Box 3015, 2601 DA, Delft, The Netherlands, Phone: +31 (0)15 2151 846, g.dibaldassarre@unesco-ihe.org

Abstract

Observation uncertainty is nowadays recognized as a serious issue undermining the reliability of hydrological studies. For instance, many recent contributions show that river flow observations are affected by errors that may reach 25% even when state-of-the-art measurement techniques are adopted. Yet, there is still little guidance by the literature on the most appropriate modelling strategies to be adopted under observation uncertainty. We carried out a series of simulation experiments and explored how the selection of appropriate model complexity can help reduce the impact of observation uncertainty. We found that model structure plays a relevant role and, in particular, a description of the relevant physical processes that come into play can effectively contribute to limit the impact of data errors and therefore significantly reduce overall uncertainty.

Keywords: hydrology, observation uncertainty, monitoring techniques, model complexity.

1. Introduction

Over the past two decades, hydrologists have devoted increasing attention to uncertainty assessment, by also focusing on the impacts of data errors on hydrological modelling and flood risk analysis (for a review, see [1-4]). Indeed, uncertainty assessment is one of the key research targets of the PUB initiative [5] of the International Association of Hydrological Sciences (IAHS). In particular, observation uncertainty has gained increasing attention, being its impact on global hydrological uncertainty more and more recognised. For instance, the term "disinformative data" was introduced by [6,7] to indicate erroneous observations that falsify the inference of water balance dynamics.

As far as river discharge observations is concerned, best practices for measuring river discharge are summarised by [8]. A first extensive review of contributions dealing with uncertainty assessment for river flow data was already proposed by [9], who concluded that uncertainty in discharge observation varies in the range 8-20% at the 95% confidence level, depending on the monitoring technique that is used.

Recently, the World Meteorological Organisation (WMO) updated the "WMO Manual for Stream Gauging" [10], which is an interesting follow up of the previous edition. In Chapter 10, which is devoted to uncertainty of discharge measurements, it is stated that for observations obtained with the velocity-area method "....it is not possible to provide absolute guidelines for making the qualitative evaluation of accuracy. As a general rule, the accuracy of most discharge measurements will be about 5 per cent, or qualitatively a good measurement. This is sometimes used as the baseline accuracy, with accuracy upgraded to excellent when measuring conditions are significantly better than average, and accuracy downgraded to fair or poor when conditions are significantly worse than average". Yet, when river flows are retrieved by using the rating

curve method, uncertainty is expected to be much more significant. For instance, errors up to 25% at the 95% confidence level were found by [11].

A relevant matter of concern is the effect of this observation uncertainty on hydrological studies. It is expected that errors of such amount can induce significant uncertainty in flood risk assessment and engineering design. For instance, it was found that accounting for precipitation errors during hydrologic model calibration significantly affects the posterior distribution of the watershed model parameters and the uncertainty of model predictions [12]. Also, a number of authors investigated the uncertainty in the design flood caused by imprecision of the river flow data [13-17]. For instance, [14,15] investigated the effect of errors in river discharge data on flood frequency analysis and proposed an inference framework which makes explicit allowance for observation errors. More specifically, it was showed that errors induced on design flood estimation by the extrapolation of the rating curve can "very substantially, indeed massively" corrupt the estimation of the design flood [15]. It was also concluded that a better characterization of the errors induced by the rating curve is needed to obtain a more precise estimation [15]. A novel methodology for evaluating the joint impact of sample variability and rating curve imprecision on design flood estimation was developed by [16]. Yet, they did not take into account the additional uncertainty in river discharge observations induced by the extrapolation of the rating curve, which is often the main source of errors in flood data. Lastly, in [17] a Bayesian framework including a multiplicative error on the rating curve was used to assess the influence of errors in discharge data on the outcomes of flood frequency analysis.

More recently, the impact of flow data observation errors on the uncertainty of design floods was investigated by [13] using different error structures. A first analysis focused on the extrapolation error and used the error structure proposed by [15]. According to this error model, it was found that a systematic underestimation or overestimation is introduced by the

extrapolation of the rating curve [13]. A second analysis was based on the assumption that observation errors are Gaussian with zero mean and standard deviation proportional to the true river discharge and it was showed that the effect of this type of error, although conservative for design purposes, may be relevant. Finally, a third analysis was performed by using the results of [11] for the characterization of the global uncertainty in river flow data. The results of this study showed that the design flood estimation may be seriously affected by errors in flood data, especially when these are systematic [13].

The above summary shows that considerable research work was devoted to estimate observation uncertainty in hydrology [9-17]. However, little attention was devoted to the most appropriate modelling approaches to be adopted to reduce its impact. For instance, the literature dedicated ample room to comparing the performance of different models (lumped versus distributed, conceptual versus physically-based, e.g. [18]), but the role of model complexity in reducing the impact of data errors has remained largely unexplored. Experiences carried out in other scientific fields highlight that efficient solutions can be identified to minimise the effect of noise or bias in the data. For instance, in statistics a lot of attention is devoted to asymptotical properties of estimators to infer the probability distribution of parameters and results. Given that uncertainty is always present in hydrological modelling and will never be eliminated [19], we believe that there is the need for developing guidelines to drive the selection of the most appropriate model and calibration strategy depending on observation uncertainty and the actual information content of data.

The purpose of this paper is to provide a first contribution to identify best modelling practices in hydrology to minimise the impact of observation uncertainty. In detail, we argue that an appropriate selection of hydrological model complexity and calibration strategy can increase the robustness of hydrological studies against data errors. We test the above hypothesis by

performing an extended set of calibration/validation experiments. To this end, we used synthetic hydrological data that were corrupted by errors, emulating the uncertainty typically affecting hydrological records [9, 11, 13-15].

Our conclusions suggest that more complex modelling approaches, based on the representation of the dominant hydrological processes, can be preferable to simpler modelling schemes, which tend to be more sensitive to data errors. Secondly, our experiments show that data still maintain a significant information even after they are corrupted by errors. Thus, particular care should be used in discarding the information provided by uncertain observations.

2. Exploring the impact of data errors on rainfall-runoff models of increasing complexity

As mentioned, how to reduce the actual impact of observation uncertainty on hydrological modelling has been insufficiently investigated. As a matter of fact, intuition suggests that data errors can seriously undermine the reliability of hydrological simulations. However, little guidance exists on the best modelling practices to be followed under observation uncertainty.

On the one hand, common sense often suggests that simpler models, which are characterised by limited flexibility, are more robust and should therefore be preferred when data errors are suspected to be relevant. On the other hand, the non-linear structure of hydrological models may propagate errors in a counterintuitive fashion.

Hence, to explore the role of model complexity in reducing the impact of observation uncertainty, we performed extensive simulation experiments, by referring to the Secchia River basin, located in Northern Italy.

2.1 The case study and the synthetic data sample

The Secchia River flows from the Apennine Mountains in south-north direction. It is a right tributary to the Po River. The contributing area is 1214 km^2 at the closure of the mountain basin, where the main stream length is about 70 km. The basin concentration time is about 12 hours. The mean annual rainfall depth ranges between 700 and more than 1500 mm/year over the basin area. Observation of mean areal hourly rainfall and temperature over the basin, as well as river flows at the closure of the mountain basin, were collected for the year 1972. Rainfall was measured by using 5 tipping bucket raingauges displaced over the basin area, while river flows were measured through the rating curve method.

A synthetic hourly rainfall, temperature and river flow data-base, covering a 55-year observation period was generated. More details can be found in [20].

Synthetic rainfall data, for the 5 raingauges mentioned above, were generated using the generalized multivariate Neyman-Scott rectangular pulses model [21] that was calibrated using the observed data. Mean areal rainfall was then computed as a weighted sum of the rainfall in each raingauge. Synthetic hourly temperature data were generated by applying a fractionally differenced ARIMA model (FARIMA, [22]). A mean areal value for temperature was obtained by rescaling the synthetic observations to the mean altitude of the basin area, by adopting a standard temperature gradient.

Synthetic river flow data were generated by using the previously generated synthetic rainfall and temperature records as input to the lumped rainfall-runoff model ADM [23]. The ADM model is a nine-parameter lumped conceptual scheme that is derived from the Xinanjiang model [24] and it is based upon the same concept of probability distributed soil moisture storage capacity. The model is divided into two main blocks. The first block represents the water balance at soil level, that is, the balance between the moisture content and the incoming (precipitation)

and outgoing (evapotranspiration, surface runoff, interflow and baseflow) water flows. The second block represents the transfer of runoff production to the basin outlet which takes place in two distinct stages. The first represents the flow along the hillslopes towards the channel network while the second represents the flow along the channel network towards the basin outlet. Evapotranspiration is computed via the radiation method that yields estimates of the hourly evapotranspirated water that are used as input variables to the ADM model. The radiation method requires estimates of the mean monthly temperature, averaged over the basin, that is computed by using long term records of historical temperature data. ADM was calibrated against historical data obtaining a Nash-Sutcliffe efficiency of 0.81 in validation [25].

2.2 The calibration and validation experiment

To investigate the role of model structure to limit or amplify the impact of data errors, we considered 3 rainfall-runoff models of increasing complexity. The first model is the linear reservoir which counts one parameter only, namely, the constant k [t] of the bottom discharge. The second model is HYMOD, a 5-parameter conceptual rainfall-runoff model that was introduced by [26]. HYMOD consists of a relatively simple rainfall excess model that is connected with two series of linear reservoirs: three identical reservoirs for the quick response and a single reservoir for the slow response. The third model is ADM, namely, the “true” model that was used to generate the synthetic river flow data. For all models evapotranspiration is accounted for in the same way as for ADM, namely, by using the radiation method.

The three models were calibrated 10 times by using different 5-year long records of rainfall, temperature and river flow, therefore using the first 50 years of the synthetic data record. It is interesting to note that these are “true” data, with no errors. Calibration for all three models was performed by using the Genoud optimisation algorithm [26]. Validation after each

calibration experiment was performed over the 5-year period extended during the years 51-55 of the synthetic record. Table 1 shows the mean Nash efficiency E_1 in validation computed over the 10 different calibration experiments, for each model. In order to better inspect the reliability of the simulation of flood flows, Nash efficiency was computed for river flows greater than the mean value (about 16 m^3/s) and greater than 50 m^3/s. These latter efficiencies are indicated with the symbols E_2 and E_3, respectively, in Table 1.

To investigate the impact of data errors on model performance, calibration/validation experiments were repeated by referring to corrupted river flow and rainfall data. Corruption was obtained by introducing 4 types of data errors, which are described here below.

- Corruption 1 – Random error in river flow data: it was obtained by multiplying each river flow observation by a random outcome from a uniform probability distribution in the range 0.8-1.2. The magnitude of the introduced error is large, but comparable with what was detected by previous analyses with respect to state of the art monitoring techniques [9,11,13].
- Corruption 2 – Systematic bias in flood flow data: it was obtained by multiplying the river flows greater than 150 m^3/s by 1.25. This is a relatively large, but not unreasonable, error that might be originated by extrapolation errors in the rating curve [14,15].
- Corruption 3 – Rainfall error: it was obtained by introducing a random error in the raingauge weights that are used to estimate mean areal rainfall. In detail, the weight of each raingauge was randomly picked up, at each time step, in the range ±20% of its optimal value according to a uniform probability distribution. The obtained weights were rescaled so that their cumulative sum was equal to one.
- Corruption 4 – Correlated noise in river flow data: it was obtained by multiplying each river flow observation by a correlated noise generated by using an autoregressive first-order linear stochastic process, with autoregressive coefficient equal to 0.5. The resulting noise has a unit

mean and a standard deviation of 0.1, namely, statistics that are consistent with errors that are typically encountered in real world data [9,11,13].

The efficiencies E_1, E_2 and E_3 obtained by analysing the corrupted data are given in Table 1. A graphical representation of the results, which refers to one single calibration/validation experiment after data corruption 2, is given in Figure 1. The left panels show scatterplots of simulated (in validation mode) river flows by the three models against the corresponding true values. The right panels provide a visual assessment of the impact of data corruption, by showing the progress, for increasing river flows, of the absolute difference w [m^3/s] between simulated river flows by the three models before and after data corruption. The plots in the right panels were smoothed with a moving average window encompassing 25 subsequent data points.

2.3 Inspecting the performances of spectral calibration

Spectral calibration of hydrological models [27,28] is a useful solution to fit process behaviours at selected time scales. Basically, spectral calibration is performed by comparing the periodogram of observed and simulated time series. In detail, the likelihood function to be maximised was originally proposed by [29] for stationary processes and is given by:

$$L(\theta) = \exp\left[-\sum_{j=1}^{N/2} \left\{ \log\left[f_M(\lambda_j, \theta) + f_e(\lambda_j) \right] + \frac{J(\lambda_j)}{f_M(\lambda_j, \theta) + f_e(\lambda_j)} \right\} \right] \qquad (1)$$

where θ is the parameter vector to be calibrated, $\lambda_j = 2\pi j/N$ are the Fourier frequencies, J is the periodogram of the N observed data, f_M is the spectral density of the model output that depends on π, and f_e is the spectral density of the model error. The above likelihood is conditioned by the assumption of independence between f_M and f_e [27].

To compute the objective function, the observed record is decomposed in the sum of sine and cosine waves with different frequency and the amplitude of each wave is matched with the corresponding one computed on the simulated record. The parameter set which ensures the best match is finally retained. If all the frequencies are matched, then the calibration produces similar results as the classical least squares procedure. However, a proper selection of the frequencies to be matched allows one to give more stress on short rather than long term properties of the observed record. When disinformative observations are suspected to be present at local scale, one possible option could be to optimise the long term properties of the series only, being the short ones possibly affected by local errors. Therefore, spectral calibration can be classified as a "selective" calibration strategy, where focus on different frequencies of the spectrum can be obtained by simply changing a routine parameter. However, it is not clear to what extent such a solution can be effective to weigh out disinformative data.

To check the effectiveness of spectral calibration, the calibration/validation experiments described in Section 2.2 were repeated by using the Whittle's likelihood as objective function, in which the first 5% of the frequencies were not considered. The optimal number of frequencies to neglect should be identified by considering the type of error that is likely to be present. The results of the calibration/validation experiment with the spectral likelihood are given in Table 1.

3 Results

Table 1 provides a comprehensive overview of the impact of data uncertainty on model results depending on model complexity. As a premise, it is interesting to stress that ADM is the model that was used to generate the synthetic data set and therefore represents the "perfect model". It is also important to highlight that validation, for each model, was performed against the "error-free data". Namely, the river flows used to validate the models were not corrupted. By

looking at the indexes and graphs presented in Table 1 and Figure 1 the following conclusions can be drawn.

- The ADM model is capable of reproducing the "error-free data" reliably (Figure 1, top left panel). The very limited validation uncertainty (Nash efficiency close to 1) can be attributed to the fact that the model was calibrated with a relatively limited sample size (5-year record), which induces some parameter uncertainty. This outcome confirms the appropriateness of the calibration/validation procedure.
- By reducing model complexity a correspondingly increasing model structural uncertainty is introduced (Figure 1, left middle and left bottom panel), which is due to progressive incapability to simulate the dominant processes and therefore the whole range of river flows. The fact the low flows dominate in number of data points (the average river flow is about 16 m^3/s) justifies the fact that the failure is particularly noticeable for the higher flows (see also the values of the efficiencies E_2 and E_3 in Table 1).
- A first relevant conclusion is that data corruption alone, although representative of the typical errors affecting hydrological observations [9,11,13], does not significantly affect the model performance in validation mode. In contrast, model structural uncertainty is responsible for a significant decrease of model performance. Actually, it brings the values of the Nash efficiencies far from unity, therefore closer to values that are typically found in practical applications. Although this result depends on the type of data corruption that is introduced, it does suggest that in many practical applications the inefficient model structure may be a major limiting factor, more than the presence of disinformative observations.
- A second relevant conclusion is that the impact of data errors depends on model complexity. This results is not well highlighted by the mean Nash efficiencies presented

in Table 1, but becomes clearer if one looks at the mean absolute difference w [m^3/s] computed over all calibration experiments. In fact, it amounts to 1.39 m^3/s, 1.56 m^3/s and 2.00 m^3/s for ADM, Hymod and linear reservoir, respectively. Moreover, a further demonstration is given by Figure 1, which refers to one calibration experiment after corruption 2. By comparing the right middle and right bottom panel with the results that refer to the "perfect model" (Figure 1, right top panel), it is clear that the impact of data error increases for decreasing model complexity. In fact, the absolute perturbation w increases for the HYMOD model and the linear reservoir. This result suggests that the use of simpler models may not help reduce the impact of data errors.

- A third relevant conclusion refers to the use of the selective calibration (obtained by discarding the highest 5% frequencies of the spectrum). While it gives similar results for the ADM model with respect to using least squares, worse efficiencies are obtained for the two less complex models, especially in the simulation of high flow conditions. Therefore selective calibration, i.e. discarding the highest frequencies that are most affected by local errors, does not help reduce the impact of data errors. This result, although still referred to the specific case study and data corruption, suggests that particular care should be used in discarding data that are supposed to be disinformative. Actually, in this case each data point brings an information, even if corrupted by an error, and the results show that neglecting it, in the attempt to cancel the error out, is not an efficient strategy. It is important to note that this is a drawback of selective calibration and not spectral calibration in general, which proved to be equivalent to least squares when all the frequencies are used.

In summary, the message that emerges from the results is that, if measurements are made with state-of-the-art techniques, observation uncertainty alone has a reduced impact with respect

to model structural uncertainty, which is also affecting the impact of other errors with a feedback effect. Therefore, improving the model structure is definitely an important step forward to reduce hydrological uncertainty and the use of simpler models should not be justified by the suspected presence of (even relevant) data errors.

The above results indicate that the sensitivity of hydrological models to observation uncertainty should be carefully examined to identify the optimal model complexity, because the actual results may contradict intuition and common sense.

It must be noted that these results are case-study dependent and therefore, in principle, not general. Moreover, different outcomes may be obtained with different data corruptions, although we note that all the different error structures used here led to similar results.

4. Discussion and Conclusions

4.1 Ways forward to reduce observation uncertainty

As mentioned, uncertainty in river flow data is much more relevant when the stage-discharge rating curve is extrapolated beyond the measurement range used for its derivation. Some authors have proposed the use of hydraulic models to reduce the inaccuracies due to the extrapolation of rating curves [27-28]. Indeed, [28] showed that the indirect observations of discharges beyond the measurement range should better rely on a physically based models, instead of traditional extrapolation approaches based on analytical black-box relationships. It should be noted hydraulic studies of river reaches are an increasingly attractive option today in view of the broad availability of topographic data and model codes [29,30]. Such studies may help to obtain more reliable stage-discharge relationships in the extrapolation zone. A possible operational strategy could be to use the stage-discharge measurements to calibrate a hydraulic model and then to use the model to extrapolate the rating curve [28].

However, it must be said that the uncertainty of the hydraulic model, which is calibrated using ordinary flow data and then used to simulate extremely high flow conditions, is often not negligible and therefore needs to be carefully evaluated. For instance, a number of studies [e.g. 31-33] have shown that the effective roughness coefficients may be different when evaluated for different flow conditions. It is then recommended to constrain the model uncertainty by using some direct observations during high water conditions. To this end, the current proliferation of remote sensing data [34-37], which has led to a sudden shift from a data-sparse to a data-rich environment for hydraulic modelling of floods [29] can be extremely valuable.

As mentioned, observation errors reduce the information content of data, to the extent that a part of the data set may become “disinformative” [6,7]. The results presented in this paper prove that particular care should be used to discard observations that may still contain some information, although corrupted. A potential solution could be to reduce the weight that is attributed to suspicious observations during the calibration phase.

Actually, observation errors are anomalies in the related time series, which in some cases can be easily identified, especially if they are locally occurring. For instance, [38,39] dealt with outlier detection in hydrological data. However, in many cases it is indeed difficult to detect anomalies in observed hydrological patterns. Modelling exercises can be an interesting opportunity to this end. In fact, the inability of a model to fit data sets (or portions of them) could be an indicator of the presence of anomalies therefore suggesting the opportunity to reduce the weight of such observations during calibration. Therefore models can play an important role to identify disinformative data sets, even if one should be conscious that the model might assume a non behavioural structure in the attempt to fit anomalous behaviours of the data. Sensitivity analysis and comparison of results may help identify the best option.

4.2 The role of model structure in limiting the impact of data errors

Observation uncertainty is increasingly recognized to significantly affect hydrological modelling [9-17, 40-42]. In the previous section, we argued that observation uncertainty can be reduced by using advanced methods offered, for instance, by the recent progresses in hydraulic modelling and remote sensing data, which can potentially open interesting perspectives. Also, simulation experiments, similar to the one herein presented, may help select appropriate modelling strategies and reduce the impact of observation uncertainty.

This paper described a numerical study attempting to decipher the impact of data errors on the results of hydrological models of increasing complexity. We referred to the case of a river basin located in Northern Italy, for which we generated an extended synthetic data set, which was subsequently corrupted by introducing different types of errors. This procedure allowed us to inspect the uncertainty that is introduced by model inadequacy and data corruption.

The result showed that, if measurements are made following state-of-the-art techniques [10,11], observation uncertainty has a limited impact, with respect to model structural uncertainty, on the outcomes of hydrological models. This result is particularly interesting as the above simplifications of the model structure are believed to be less impacting with respect to the actual simplifications introduced by hydrological models (with respect to reality). Moreover, model structural uncertainty also induces a feedback on the impact of data errors, which appears to be more significant for simpler models. Hence, improving the model structure via process understanding appears to be a crucial step forward to reduce hydrological uncertainty.

Although these results depend on the specific case study and types of data corruption, they indicated that the sensitivity of hydrological models to observation uncertainty should be carefully examined to identify optimal model complexity and efficient strategies to limit the

impact of data errors, because the actual results may contradict intuition and common sense. For instance, the use of simpler models cannot be justified by the presence of data errors only.

Lastly, the results showed that particular care should be taken in discarding the information content of uncertain observations. In hydrological modeling any information is important and the presence of data errors does not necessarily limit the usefulness of observed record.

Acknowledgements

The authors are extremely grateful to three anonymous referees that contributed to improve the earlier version of the manuscript. AM was partially supported by the Italian government through the grant "Uncertainty estimation for precipitation and river discharge data. Effects on water resources planning and flood risk management", while GDB was partially supported by the EC FP7 Project KULTURisk: "Knowledge-based approach to develop a cULTUre of Risk prevention" (Grant Agreement Number: 265280).

References

[1] Montanari, A.: What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology, Hydrol. Proc., 21, 841–845, 2007.

[2] Montanari, A., Uncertainty of hydrological predictions, in Treatise on Water Science, edited by Peter A. Wilderer, Elsevier., 2011.

[3] Montanari, A., Shoemaker, C.A., van de Giesen, N., Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, Water Resources Research. vol. 45, doi:10.1029/2009WR008471, 2009.

[4] Beven, K.J. Environmental Modelling: An Uncertain Future?, Routledge, London, 2008.

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

[6] Beven, K. J. and Westerberg, I. K. On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol. Process., 25, 1676-1680, 2011.

[7] Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205-2227, doi:10.5194/hess-15-2205-2011, 2011.

[8] European ISO EN Rule 748, Measurement of liquid flow in open channels - velocity-area methods, Reference number ISO 748:1997 (E), International Standard, 1997.

[9] Pelletier, M.P.: Uncertainties in the determination of river discharge: a literature review, Can. J. Civ. Eng., 15, 834–850, 1987.

[10] World Meteorological Organisation: Guide to Hydrological Practice, WMO Pubbl. no. 168, 1994.

[11] Di Baldassarre G., Montanari, A. Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Sys. Sci., 13, 913–921, 2009.

[12] Vrugt, J.A., C.J.F. ter Braak, M.P. Clark, J.M. Hyman, and B.A. Robinson. Treatment of input uncertainty in hydrological modeling: doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, 2008.

[13] Di Baldassarre G., F. Laio, A. Montanari. Effect of observation errors on the uncertainty of design floods, Physics and Chemistry of the Earth, doi; 10/1016/j.pce.2011.05.001.

[14] Kuczera, G. Uncorrelated rating curve error in flood frequency inference, Water Resour. Res., 28 (1), 183-188, 1992.

[15] Kuczera, G. Correlated rating curve error in flood frequency inference, Water Resour. Res., 32(7), 2119-2127, 1996.

[16] Petersen-Øverleir, A., Reitan, T. Accounting for rating curve imprecision in flood frequency analysis using likelihood-based methods, Journal of Hydrology, 366, 89–100, 2009.

[17] Lang, M., Pobanz, K., Renard, B., Renouf, E., Sauquet, E. Extrapolation of rating curves by hydraulic modeling, with application to flood frequency analysis, Hydrol. Sci. J., 55(6), 883-897, 2010.

[18] Refsgaard J.C. Parametrisation, calibration and validation of distributed hydrological models. Journal of Hydrology, 198, 69-97, 1997.

[19] Koutsoyiannis, D., and A. Montanari. Statistical analysis of hydroclimatic time series: Uncertainty and insights, Water Resources Research, 43 (5), W05429, doi:10.1029/2006WR005592, 2007.

[20] Montanari A. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 41: W08406, DOI:10·1029/2004WR003826, 2005.

[21] Cowpertwait, P.S.P. A generalized spatial–temporal model of rainfall based on a clustered point process. Proceedings of the Royal Society of London 450, 163–175, 1995.

[22] Montanari, A., R. Rosso, and M. S. Taqqu. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation, Water Resour. Res., 33, 1035– 1044, doi:10.1029/97WR00043, 1997.

[23] Franchini, M. Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall runoff models, Hydrol. Sci. J., 41, 21– 39, 1996.

[24] Zhao, R.J., Zhang, Y.L., Fang, L.R.. The Xinanjiang model,Hydrological Forecasting Proceedings Oxford Symposium. IASH, vol. 129, 1980, pp. 351–356, 1980.

[25] Nash, J. E. and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10 (3), 282–290.

[26] Boyle, D.P., Multicriteria calibration of hydrological models, PhD Dissertation, Department of Hydrology and Water Resources, University of Arizona, 2000.

[27] Montanari, A., Toth, E., Calibration of hydrological models in the spectral domain: An opportunity for ungauged basins?, Water Resour. Res., Vol. 43, W05434, doi:10.1029/2006WR005184, 2007.

[28] Quets, J.J., G.J.M. De Lannoy, and V.R.N. Pauwels, Analysis of spectral calibration methods for precipitation-discharge processes, Hydrological Processes, 24(8), 1048-1062, 2010.

[29] Whittle, P. The analysis of multiple stationary time series. J. Roy. Statist. Soc. Ser. B. 15 (1): 125–139, 1953.

[30] Horritt, M.S., Bates, P.D, Fewtrell, T. J., Mason, D. C., Wilson, M. D. (2010). Modelling the hydraulics of the Carlisle 2005 flood event, Proceedings of the Institution of Civil Engineers, Water Management, 163, 273–281.

[31] Di Baldassarre, G., P. Claps, (2011). A hydraulic study on the applicability of flood rating curves, Hydrology Research, 42 (1), 10–19, doi: 10.2166/nh.2010.098.

[32] Bates, P.D. (2004). Remote sensing and flood inundation modelling. Hydrological Processes, 18, 2593–2597.

[33] LeFavour, G., Alsdorf, D. (2005). Water slope and discharge in the Amazon river estimated using the shuttle radar topography mission digital elevation model. Geophys. Res. Lett., 32, L17404.

[34] Romanowicz, R., Beven, K., (2003). Estimation of flood inundation probabilities as conditioned on event inundation maps. Water Resources Research 39 (3), 1073–1085.

[35] Di Baldassarre G., Schumann, G., Bates, P.D. (2009). Near real time satellite imagery to support and verify timely flood modelling, Hydrological Processes, DOI: 10.1002/hyp.7229.

[36] Di Baldassarre, G., G. Schumann, L. Brandimarte, P.D. Bates (2011). Timely Low Resolution SAR Imagery to Support Floodplain Modelling: a Case Study Review, Survey in Geophysics, 32(3), 255-269, doi: 10.1007/s10712-011-9111-9.

[37] Schumann, G., Di Baldassarre, G., Alsdorf, D.E., Bates, P.D. (2010). Near real-time flood wave approximation on large rivers from space: application to the River Po, Northern Italy. Water Resources Research, 46, W05601.

[38] Kundzewicz, Z.W., Change detection in hydrological records - a review of the methodology, Hydrological Sciences Journal, 49, 7-19, 2004.

[39] Laio, F., Allamano, P., and Claps, P.: Exploiting the information content of hydrological "outliers" for goodness-of-fit testing, Hydrol. Earth Syst. Sci., 14, 1909-1917, 2010.

[40] Kavetski, D., G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005WR004368, 2006a.

[41] Kavetski, D., G. Kuczera, and S. W. Franks. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42, W03408, doi:10.1029/2005WR004376, 2006b.

[42] McMillan H, Jackson B, Clark M, Kavetski D, Woods R. Rainfall uncertainty in hydrologic modelling: An evaluation of multiplicative error models, Journal of Hydrology, 400, 83-94, 2011.

Table and Figure Caption

Table 1. Mean Nash efficiency E_1, E_2 and E_3 in validation for the three considered models computed for the 10 different calibrations performed by using 5-year observation periods (years 1-50 of the synthetic record) and least squares as objective function. E_1, E_2 and E_3 are computed on all river flows, river flows greater than the mean value (about 16 m^3/s) and greater than 50 m^3/s, respectively. Validation was performed by simulating the last 5-year period (years 51-55) of the synthetic record. Efficiencies for uncorrupted and corrupted data are reported (see Sections 2.2 and 2.3).

	Uncorrupted			Corruption 1			Corruption 2			Corruption 3			Corruption 4		
Model	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
Linear reservoir	0.66	0.05	-1.10	0.66	0.05	-1.09	0.68	0.23	-0.56	0.66	0.07	-1.02	0.66	0.05	-1.10
Hymod	0.85	0.74	0.56	0.85	0.74	0.56	0.86	0.77	0.63	0.85	0.75	0.58	0.85	0.74	0.57
ADM	0.96	0.93	0.88	0.96	0.93	0.89	0.95	0.92	0.88	0.96	0.93	0.89	0.96	0.93	0.89
Linear reservoir (spectral)	0.60	-0.32	-2.16	0.59	-2.7’	-8.36	0.65	-0.02	-1.30	0.59	-2.64	-8.18	0.52	-0.77	-3.45
Hymod (spectral)	0.76	0.60	0.34	0.84	0.73	0.54	0.81	0.71	0.56	0.84	0.73	0.53	0.70	0.49	0.37
ADM (spectral)	0.95	0.91	0.87	0.96	0.92	0.88	0.95	0.91	0.87	0.96	0.92	0.88	0.96	0.92	0.87

Figure 1. Performance of the three models and data uncertainty for a selected calibration experiment (data corruption 2). The left panels show scatter-plots of observed versus simulated data without data corruption. The right panel show the progress, with respect to the magnitude of the true river flows, of the absolute differences between model simulation before and after data corruption.

Figure

Click here to download high resolution image