Reply to "Comment on 'A blueprint for process-based modeling of uncertain hydrological systems'" by G. Nearing

Alberto Montanari*[1] and Demetris Koutsoyiannis[2]

1. Department of Civil, Environmental, Chemical and Materials Engineering (DICAM), University of Bologna, Bologna, Italy
2. Department of Water Resources and Environmental Engineering, National Technical University of Athens, Zographou, Greece

Short title:

REPLY

* Corresponding author: A. Montanari, Department of Civil, Environmental, Chemical and Materials Engineering (DICAM), University of Bologna, Via del Risorgimento 2, IT-40136, Bologna, Italy. (alberto.montanari@unibo.it)

1. Introduction

Nearing [2014] (hereafter Nearing) has two major concerns on *Montanari and Koutsoyiannis* [2012] (hereafter MK), i.e., that the method presented by MK: (1) is not novel and (2) is not likelihood-free. He concludes that the MK "…strategy does not inherently address the underlying issues". We discuss the above major concerns in the following two sections of our reply. We also discuss several other minor remarks by Nearing in Section 4 and 5. We are grateful to Nearing for his interest in the method we proposed in MK. We consider the present reply an opportunity to further clarify several features of our approach.

2. Novelty of the method

Nearing implies that the method presented by MK is not new. This emerges in several of his statements, which we are discussing, and replying to, here below. First, after presenting his eq (1), that is a rewriting of eq. (8) in MK, Nearing states "A ubiquitous example of applying (1) with Monte Carlo integration to turn deterministic models into stochastic models is in ensemble data assimilation, where $f(Q)$ represents a Bayesian prior distribution over the current state of a dynamic system estimated as the sum of a deterministic model prediction plus random error…". Several references follow this sentence in the comment.

Actually, in none of the references cited by Nearing can we find a derivation, and neither a use, of eq. (8) that was presented in MK. The cited papers include several formulations where uncertainty is accounted for by adding a random error to a deterministic formulation. This approach is used not only in the context of data assimilation, but also in several other approaches to account for inherent uncertainty in hydrological modeling. The addition of a random error to a deterministic model is indeed the premise of the approach proposed in MK, which is resembled by our eq. (3). By relying on the above premise, the essence of the MK contribution is to analytically derive eq. (8) to prove that the probability distribution of the model output can be estimated, under appropriate assumptions, by using the probability distribution of the model error to account for model structural uncertainty and all other uncertainties that are not explicitly accounted for. In eq. (8) presented in MK $f(Q)$ is not a “Bayesian prior distribution” as Nearing implies, by comparing or identifying it with his example of “ensemble data assimilation”, but rather the probability distribution of the predictand once the prediction has been made by using all the available information.

Later on, Nearing states that “.. if we know the necessary distributions, then (1) is simply the straightforward Bayesian solution….”. This sentence as well seems to imply that Nearing is convinced that the MK approach is largely used already. We cannot agree with this statement. A Bayesian solution presupposes assuming a prior distribution for the predictand that is updated by using a likelihood function. Neither the prior nor the likelihood are included in eq. (8) derived by MK and therefore the MK approach can hardly be defined a “straightforward Bayesian solution”. It seems that Nearing is at least missing the point that MK do not use any likelihood function in their eq. (8), and therefore the latter cannot be viewed as a Bayesian solution. The likelihood is avoided by MK by using the probability distribution of the model error, which incorporates different information that nevertheless can be estimated empirically. The peculiarities of the distribution of the model error with respect to the likelihood will be further discussed in the next Section of our reply.

On the other hand, the original contribution provided by eq. (8) in MK seems to be later recognized by Nearing himself, when he writes that “… to my knowledge no previous study has actually implemented (2)”, and “*Beven and Binley* [1992] …. did not sample model error …”. In fact, we fully agree with the above considerations which confirm and highlight the novel items in the MK approach within the hydrological literature.

Indeed, we do not claim that what we are proposing is revolutionary: our theoretical analysis makes use of fundamental concepts of stochastics. However, we maintain that our methodological scheme, along with the analysis of its stochastic properties, constitutes a novel contribution to hydrology.

3. Is MK likelihood-free?

Before discussing Nearing's point that the MK approach is not likelihood-free, we believe it is necessary to provide a definition of the likelihood function. Following Nearing, we adopt the definition by *Fisher* [1922]: "*[t]he likelihood of any parameter (or set of parameters) should have any assigned value is proportional to the probability that if this were so, the totality of observations should be that observed.*" More formally, given a certain model with parameters **Θ** and given observations **Y** (which here could be thought of as a vector of past values *Q*), the likelihood is a function of the parameters **Θ** of the model proportional to the probability of those observed outcomes given those parameter values, i.e. $L(\mathbf{\Theta}|\mathbf{Y}) = p(\mathbf{Y}|\mathbf{\Theta})$, where *p* denotes probability (or, if **Y** is a vector of continuous variables, probability density). It is important to point out that (1) the likelihood is a function of the parameters **Θ** as the observations **Y** are known numbers; (2) for a specified parameter set **Θ**, the likelihood is a number; (3) for varying **Θ**, the likelihood equals a probability (or a probability density) but is not a probability distribution with respect to **Θ** (or **Y**, which in fact is a set of numbers); and (4) assuming that the model is used to predict a future (true) value *Q*, belonging to the same process as the observations **Y**, obviously the likelihood is not a function of the predictand *Q*. The last point seems trivial, but it is relevant to the discussion that follows here below.

Nearing provides several statements to claim that the probability distribution of the model error that is used by MK in eq. (8) is a likelihood, therefore concluding that MK's method is not likelihood free. For instance, he states that "…and second, that $f_{e|D}$ is a likelihood function associated with many of the issues outlined above", where $f_{e|D}$ is the probability distribution of the model error (note that Nearing explicitly recognizes the dependence of the estimated probability densities on the observed data *D*. Such dependence is implicitly recognized in MK, where we preferred to keep the notation simple). Thus, Nearing concludes that "MK's formulation does not really avoid the need to evaluate a likelihood function". We do not agree with the above conclusion for the reasons that we explain here below. For the sake of clarity, let us point out that the probability distribution of the model error *e* is indicated in MK with the symbol $f_e(Q - S(\mathbf{\Theta}, \mathbf{X})|\mathbf{\Theta}, \mathbf{X})$, where *e* is the model prediction error, *Q* is the true variable to be predicted and *S* is the output from the considered deterministic hydrological model that depends on the model parameter vector $\mathbf{\Theta}$ and input data $\mathbf{X}$.

Nearing notes that the probability distribution of the model error, which he denotes as $f_{e|D}$, can be rewritten as a probability distribution of *Q*, which he denotes with $f'_{e|D}$. Then, he concludes that "$f'_{e|D}$ (and thus the equivalent $f_{e|D}$) is a likelihood function according to Fisher's definition." We maintain that the probability distribution of the model error, according to the definition of likelihood provided above, is not a likelihood function. The simplest way to prove our assertion is to note that $f_e(Q - S(\mathbf{\Theta}, \mathbf{X})|\mathbf{\Theta}, \mathbf{X})$ depends on the value of the true variable to be predicted *Q* and therefore is different with respect to a likelihood function (see the definition of likelihood that was given above and remember that *Q* is a variable to be predicted and therefore is unknown, namely, it is not an observation).

One may counter argue that the vector of observations $\mathbf{Y}$ in the likelihood function $p(\mathbf{Y}|\mathbf{\Theta})$ as defined above is in essence equivalent to the predictand *Q* and therefore the likelihood $p(\mathbf{Y}|\mathbf{\Theta})$ is mathematically equivalent (or can be derived from) the distribution $f(Q|\mathbf{\Theta})$. This seems to be the line of thought of Nearing who (a) identifies $f_e(Q - S(\mathbf{\Theta}, \mathbf{X})|\mathbf{\Theta}, \mathbf{X})$ with $f'_{e|D}$, (b) regards the latter as probability density of *Q*, and (c) calls it likelihood. Here we note that, had we assumed an analytical form of the multivariate probability distribution of the model error (for an arbitrarily long dimensionality or a vector of errors), then from this analytical form one could indeed analytically derive $f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{\Theta})$ and hence likelihood, so that specifying the former allows one to derive the latter. However, this is not the case in MK. We did not specify any mathematical form for the multivariate probability distribution of the model error and we did not calculate the likelihood. Thus, in MK, the probability distribution of the model error $f_e(Q - S(\mathbf{\Theta}\ , \mathbf{X})|\mathbf{\Theta}, \mathbf{X})$ is just a conditional predictive distribution of e given the parameters $\mathbf{\Theta}$ and the inputs $\mathbf{X}$. It is not a likelihood per se. Clearly, it is a marginal distribution for one variable e (or Q) and not a multivariate distribution, whereas, as already explained, the observations $\mathbf{Y}$ form a vector and $f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{\Theta})$ is a multivariate distribution. Because of complex dependencies, the multivariate distribution cannot be derived from marginal distributions in a trivial manner (e.g. as a product of marginal densities).

Finally, let us note another reason why the probability distribution of the model error cannot be a likelihood, which is that the likelihood is a probability, and not a probability distribution (this is quite well understood in the scientific community and there is extensive information on this, as one can confirm by searching the web for "likelihood is not a probability distribution").

We hope that our above explanations fully justify our claim that we did not calculate or evaluate any likelihood in MK. We did not make any attempt to formulate an analytical form for the likelihood or to estimate it empirically (and neither *Montanari and Brath* [2004] did) because we did not see the reason to do that. Furthermore, we note that the probability distribution of the model error can be derived by using any type of information, including observed data but also soft information and/or expert knowledge. This is another striking difference with respect to a likelihood function that is computed over observed data.

Another interesting remark is raised by Nearing when he points out that MK used a likelihood to estimate the probability distribution of the model parameters. In fact, he notes that "…MK evaluate $f_{D|\Theta}$ in their application of DREAM [*Vrugt and Robinson,* 2007] to estimate $f_{\Theta|D}$".

In principle Nearing is correct in this point, but we thought we had made it clear in MK that the proposed method is likelihood free in our eq. (8) of MK. We never stated that the use of the likelihood should be (or should not be) avoided in the whole calibration/validation/simulation process. Rather, we pointed out that the use of a formal likelihood is problematic when assessing the uncertainty of the predictand. Therefore, in MK it was decided to use a surrogate of a likelihood function (sum of squared errors) for parameter calibration but such likelihood was dismissed when estimating the uncertainty of the model predictions. It is relevant to note that such a procedure is not inconsistent. Assumptions that are acceptable for parameter estimation may be no more justified when estimating uncertainty, because of the different impact that the same assumption may have on different procedures of statistical inference.

Nearing's states that "… the purpose of this comment is to point out that, although it is possible to construct methods that avoid likelihood evaluation, this objective is something of a red herring and does not address the fundamental issues". Therefore, he objects that using the probability distribution of the model error does not simplify the problem with respect to computing the likelihood. This latter view is also supported by Nearing by noticing in his comment that MK do not take into account possible dependency and non-stationarity in the model error, as well as the dependence of the model error on model parameters and input data. Therefore, he concludes by stating that the MK "…strategy does not inherently address the underlying issues".

Actually, the above problems are extensively discussed in MK. We would like to point out once again that our scheme, being very general, offers the grounds for resolving the above limitations mentioned by Nearing. Error dependence and possible non-stationarity can be accounted for in MK by properly defining the probability distribution of the model error, provided enough information is available. In fact, by proposing to estimate the probability distribution of the model error empirically, MK identify a way to make a significant step forward to reach the target. For an additional discussion on this issue the interested reader is invited to refer to *Sikorska et al*. (2014). Likewise, MK already pointed out that the joint probability distributions of the model error, input data and parameters can be in principle estimated by relying again on an empirical (although much more computer intensive) approach (see the discussion in Sections 5 and 6.4 in MK).

To conclude this Section, we recognize that Nearing is certainly correct in pointing out that MK did not resolve all the problems related to uncertainty estimation. We never claimed that we achieved such ambitious result, given that we explicitly recognize in MK that some significant challenges still stand. However, we believe that our approach represents a significant step forward, in that it allows one to avoid significant problems related to likelihood identification and estimation — and here we disagree with Nearing's main point.

4. Is MK conditioning uncertainty estimation on evidence?

Nearing seems to imply that MK do not use a fully Bayesian approach and therefore do not efficiently compute the posterior distribution of the predictand basing on the available observations. To clarify this issue it is useful to define the event which marks the difference between the prior (to the event) and posterior inference. In our interpretation such event is given by the observation of data that could be used to condition uncertainty assessment, namely, the data that in a Bayesian approach would be used to compute the likelihood function in order to update a prior distribution.

This gives us the opportunity to note that the method proposed by MK does not necessarily need observed data to be applied, given that the probability distributions of the input data and parameters, as well as that of the model error, may be estimated by using soft information and/or expert knowledge. However, if observed data are available (the evidence), the MK method does allow one to condition uncertainty estimation for the model output on the observations themselves. In fact, in their applications presented in the paper, MK estimate the probability distribution of the model error based on observed data, which are used to compute the error itself for the sake of inferring its distribution. Such conditioning procedure can also be applied in real time in a data assimilation context, by updating the probability distribution of the model error, and perhaps the probability distributions of input data and parameters, as new observations become available. From a technical point of view, MK also proved that the conditioning was successful in the developed case studies, as the coverage probabilities plots shown in figures 5 and 9 in MK clearly show. Therefore, we conclude that the updated probability distribution that follows the acquisition of new data (or, generally speaking, follows any kind of available information) is indeed estimated by MK.

5. Reply to Nearing's minor comments

In addition to the above major issues, Nearing offers some minor criticism on MK. First, Nearing states that "*Montanari and Koutsoyiannis* [2012] … offer an excellent discussion of the fundamental role of epistemic uncertainty in hydrologic modeling". And, later on, he writes "I will show that they have not actually avoided likelihood evaluation, but that their method nevertheless offers very meaningful insight into the fundamental issues associated with applying probability theory to estimate epistemic uncertainty".

We think it is useful to make clear that MK refer to uncertainty in general and not to epistemic uncertainty in particular. In MK, we provided a discussion on the role of epistemic uncertainty to reply to questions and criticism by the referees, but we believe that it is misleading to introduce a strict classification of uncertainty (we see in the literature many attempts to split uncertainty in categories like epistemic, non-epistemic, random, non-random, and many others). We prefer to avoid such a classification because these components cannot be separated in practice. We prefer to associate uncertainty to unpredictability, without attempting to make any decomposition. It is clear that uncertainty can be potentially reduced, but we believe it is inappropriate to rigorously attempt to separate reducible from irreducible uncertainty. Uncertainty decomposition is subjective, uncertain and unnecessary.

We do not understand Nearing's comment "Isolated application of (1) is not particularly useful because it requires a priori knowledge of all model components (e.g., e, $\mathbf{\Theta}$ and $\mathbf{X}$)". Of course any model application requires the knowledge of all model components. We do not see any problem in recognizing that any modeling exercise should be based on information and what is proposed in MK does not require more information with respect to alternative approaches. We only need a model and data (or an alternative information), like any application in hydrology.

After presenting eq. (3) in the comment Nearing states that "The only difference between (2) and (3) is that (3) does not require model error to be additive". Actually, we do not see the need to introduce or use his eq. (3). The use of an additive error in MK results in a mathematically consistent equation (eq. (3) in MK). Actually, the usage of the error e is not removed in Nearing's eq. (3) as is clear in the first tem of the right hand side. Furthermore, we believe there is a formal error in eq. (3) as the right-hand side is clearly a mathematical function of the error e , while the left-hand side is a function of Q.

6. Conclusion

It has been a common ground that, from a Bayesian perspective, any attempt to fit a model should necessarily involve its likelihood. Indeed, if one adopted and extended the definition of likelihood, by including empirical estimation and informal approaches for estimation based on expert knowledge, we would agree that any uncertainty assessment method that is conditioned by information is not likelihood free. This is a fully motivated and scientifically sound view which we are not questioning, but our aim is not to discuss the advantages of Bayesian methods. We are not interested in classifying our method as Bayesian or not Bayesian. However, we think it is interesting to point out that in our derivations we did not follow a standard Bayesian approach (although perhaps this would be possible, yet not actually provided by Nearing). Our motivation for studying an approach that avoids the analytical specification and numerical computation of a formal likelihood function for hydrological models is only the will to propose a viable solution to address real world problems. We believe that there is an urgent technical need to provide reliable assessments of uncertainty in hydrology and therefore we are making an effort to propose a solution that in our opinion is theoretically justified, practical and susceptible to further simplification [*Sikorska et al*., 2014].

A criticism that can be cast to our method, which is implied by Nearing, is that defining a probability distribution for the model error is even more complicated than computing the likelihood function. We would agree with this criticism if we were compelled to use a formal analytical approach to uncertainty estimation. Conversely, if an empirical, Monte Carlo, approach is adopted, as we propose, then estimating the probability distribution of the model error is simpler than estimating a likelihood. In fact, likelihood computation requires an analytical description of the statistical features of the error, and in particular its dependence properties and temporal variability of statistics. This is frequently achieved by specifying a model for the error itself, which implies relevant problems related to the analytical interpretation of error features. This complication is avoided if the probability distribution of the model error is evaluated conditionally on the deterministic model prediction (which is time varying). That can be empirically estimated in a simple manner if enough data are available (for an example of application, see *Sikorska et al*. [2014]).

Our approach was proposed after numerous attempts to estimate and check uncertainty assessment in practice: at the end we gained the convincement that our method deserves to be known. Uncertainty assessment is a tremendously important issue in hydrological practice, as we all know: we need to make an effort to systematize the underlying theory without being trapped in stereotypical classifications.

Acknowledgements

We are grateful to G. Nearing for providing an opportunity to further discuss and better clarify the MK approach. We are also grateful to D. Del Giudice, A. Scheiddeger, C. Albert, J. Rieckermann and P. Reichert (Eawag Zürich) for providing a post-publication review of MK. We are also thankful to two anonymous reviewer of this reply. The work presented here has been carried out in the framework of the Panta Rhei research initiative of the International Association of Hydrological Sciences.

References

Fisher, R.A. (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, *222*, 309-368, doi:http://dx.doi.org/310.1098/rsta.1922.0009.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resources Research*, *40*, W01106; doi: http://dx.doi.org/01110.01029/02003WR002540.

Montanari, A., and D. Koutsoyiannis (2012), A blueprint for process-based modeling of uncertain hydrological systems, *Water Resources Research*, *48*, W09555; doi: http://dx.doi.org/09510.01029/02011WR011412.

Nearing, G. (2013), Comment on “A blueprint for process-based modeling of uncertain hydrological systems” by A Montanari and D. Koutsoyiannis, under review.

Sikorska, A., A. Montanari, and D. Koutsoyiannis (2014), Estimating the uncertainty of hydrological predictions through data-driven resampling techniques, ASCE Journal of Hydrological Engineering, in press.