AGU PUBLICATIONS

Water Resources Research

COMMENT

10.1002/2013WR014812

This article is a comment on *Montanari and Koutsoyiannis* [2012], doi:10.1029/2011WR011412.

Correspondence to:
G. Nearing,
grey.s.nearing@nasa.gov

Citation:
Nearing, G. (2014), Comment on "A blueprint for process-based modeling of uncertain hydrological systems" by Alberto Montanari and Demetris Koutsoyiannis, *Water Resour. Res.*, *50*, 6260–6263, doi:10.1002/2013WR014812.

Received 27 SEP 2013
Accepted 18 JUN 2014
Accepted article online 23 JUN 2014
Published online 24 JUL 2014

Comment on "A blueprint for process-based modeling of uncertain hydrological systems" by Alberto Montanari and Demetris Koutsoyiannis

Grey Nearing[1]

[1]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Montanari and Koutsoyiannis [2012] (hereafter referred to as MK) offer an excellent discussion of the fundamental role of epistemic uncertainty in hydrologic modeling. They point out that epistemic uncertainty is unavoidable, and that "one of the most comprehensive, elegant and complete ways of dealing with uncertainty is provided by the theory of probability." In fact, we know from *Cox*'s [1946] theorem that probability theory provides the *only* method for quantifying epistemic uncertainty that is consistent with certain basic principles of logic [*Howson and Urbach*, 1989; *Rathmanner and Hutter*, 2011; *Van Horn*, 2003].

MK claim to have avoided the need to evaluate a likelihood function during Monte Carlo Bayesian uncertainty propagation by estimating and sampling a distribution over model error directly. 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. I argue that there are three such issues: *subjectivity*, *nonstationarity*, and *dimensionality* (explained presently), and 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. Because rigorous efforts to estimate uncertainty will necessarily be built on *Kolmogorov*'s [1956] axioms, it is important that we understand and are explicit about the core issues.

MK propose to "estimate[e] the probability distribution of the output from a process-based (deterministic) hydrological model." Their main result is that the probability of the true value of a hydrologic variable to be predicted, Q, can be estimated from deterministic model S as (MK's equation (8)):

$$f_Q(Q)=\int\int f_e(e|\boldsymbol{\Theta},\boldsymbol{X})f_\Theta(\boldsymbol{\Theta})f_X(\boldsymbol{X})d\boldsymbol{\Theta}d\boldsymbol{X}, \tag{1}$$

where $\boldsymbol{\Theta}$ are model parameters, $\boldsymbol{X}$ are model inputs, $e=Q-S(\boldsymbol{\Theta},\boldsymbol{X})$ are model residuals, and f_* are probability distributions. Notice that the right-hand side gives a distribution over e, and equality results from the fact that the mapping from Q to e is bijective for any given $\boldsymbol{\Theta}$ and $\boldsymbol{X}$. MK use Monte Carlo integration to marginalize over $\boldsymbol{\Theta}$ and $\boldsymbol{X}$.

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 [*Evensen*, 2003; *Nearing et al.*, 2012; *Reichle et al.*, 2008; *Vrugt et al.*, 2006, 2005]. Isolated application of (1) is not particularly useful because it requires a priori knowledge of all model components (e.g., e, $\boldsymbol{\Theta}$, and $\boldsymbol{X}$); if we know the necessary distributions, then (1) is simply the chain rule. Data assimilation deals with this by conditioning f_Q directly on some observation data $\boldsymbol{D}$ using Bayes' law. (Incidentally, MK state that "[l]ikelihood computation might . . . be avoided by using data assimilation," however, it should be noted that data assimilation is a Bayesian method and therefore requires a likelihood function in the form of an observation operator [*van Leeuwen*, 2010; *Wikle and Berliner*, 2007] that suffers from the standard issues [*Snyder et al.*, 2008; *Zupanski*, 2005].)

Another way to condition on data is to use $f_{\Theta|D}(\boldsymbol{\Theta}|\boldsymbol{D})$ and $f_{e|D}(e|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{D})$ to obtain:

$$f_{Q|D}(Q|\boldsymbol{D})=\int\int f_{e|D}(e|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{D})f_{\Theta|D}(\boldsymbol{\Theta}|\boldsymbol{D})f_X(\boldsymbol{X})d\boldsymbol{\Theta}d\boldsymbol{X}. \tag{2}$$

MK notate their method as (1) but actually employ (2). Although a number of studies advocate considering parameter uncertainty, forcing uncertainty, and simulator discrepancy as the major components in an uncertainty forecasting paradigm [e.g., *Liu and Gupta*, 2007; *Wagener and Gupta*, 2005], to my knowledge, no previous study has actually implemented (2). For example, *Wilkinson et al*. [2011] sampled $f_{e|D}$ for a single realization of parameters and inputs, and *Harrison* [2007] sampled $f_{e|D}$ and marginalized over $f_{\Theta|D}$, but did not consider input uncertainty explicitly. *Beven and Binley* [1992] marginalized over $f_{\Theta|D}$, but did not sample model error or explicitly consider input uncertainty.

MK note problems with many existing methods for estimating total uncertainty over deterministic model predictions, most of which center around the use of likelihood function. Likelihood was defined by *Fisher* [1922] as "[t]he likelihood of any [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." In the context of (2), a rigorous likelihood function like $f_{D|\Theta}(\boldsymbol{D}|\boldsymbol{\Theta})$ would be a convolution of a model error distribution and an observation error distribution. MK note that likelihood functions are difficult to develop because "the complex structure of the model error . . . makes its statistical description complicated," and that they are often "based on assumptions that may be restrictive in some practical applications, like . . . independence for the model error." They also note that likelihood functions cause problems during implementation because "the likelihood is usually estimated in calibration, . . . but is used to assess uncertainty of out-of-sample predictions." I would also highlight another issue: when likelihood is sampled using large data sets in a Monte Carlo context, there is, due to the *curse of dimensionality* [*Bellman*, 2003], propensity for sample collapse—a large majority of samples are assigned negligible probabilities [e.g., *Snyder et al*., 2008]. Therefore, we seem to have three complications related to the use of likelihood functions: (1) understanding the statistical properties of model error (*subjectivity* or the *estimation* problem), (2) the possibility that these statistical properties change over time (*nonstationarity*), and (3) the fact that it is difficult to sample high-dimensional probability distributions (*dimensionality*).

MK advocate (1) (actually (2)) as avoiding many of these issues by not requiring evaluation of a likelihood function. I have two issues with this claim. First, MK evaluate $f_{D|\Theta}$ in their application of DREAM [*Vrugt and Robinson*, 2007] to estimate $f_{\Theta|D}$, and second, that $f_{e|D}$ is a likelihood function associated with many of the issues outlined above. Because the mapping from Q to e is bijective, there exists an $f'_{e|D}$ such that $f'_{e|D}(Q|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{D})=f_{e|D}(e|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{D})$, and (2) can be rewritten as:

$$f_{Q|D}(Q|\boldsymbol{D})=\int\int f'_{e|D}(Q|\boldsymbol{\Theta},\boldsymbol{X},\boldsymbol{D})f_{\Theta|D}(\boldsymbol{\Theta}|\boldsymbol{D})f_X(\boldsymbol{X})d\boldsymbol{\Theta}d\boldsymbol{X}. \tag{3}$$

The only difference between (2) and (3) is that (3) does not require model error to be additive, although it applies isomorphically in that case. $f'_{e|D}$ (and thus, the equivalent $f_{e|D}$) is a likelihood function according to Fisher's definition. It is worth pointing out that although MK call Q "the true value of the hydrologic variable to be predicted," they actually estimate a distribution over the discrepancy between model predictions and observations, and thus their f_e implicitly accounts for both model and observation error.

Leaving aside the fact that MK use $f_{\Theta|D}$ in DREAM (this has interesting implications, and we will come back to it), MK's use of a likelihood function in the form of $f_{e|D}$ differs from the methods they criticize in that they do not use it to compute the likelihood of any set of observations—instead they sample $f_{e|D}$ directly. By not evaluating $f_{e|D}$, they avoid the problem of sample collapse—all of their samples of e are iid. Further, $f_{e|D}$ is derived directly from data and avoids many of the strong assumptions that are present in the likelihood functions used by a number of previous studies. Similarly, by conditioning model error on currently available information (i.e., $\boldsymbol{X}$ and $\boldsymbol{\Theta}$) MK account for some of the heteroscedasticity in e. In theory, one could imagine estimating an asymptotically correct, nonparametric $f_{e|D}$ purely empirically (as a kernel density function), and therefore nearly completely alleviating the need for parametric assumptions. In practice, however, this is impossible due to the curse of dimensionality, and assumptions are necessary. For example, MK assume e to be independent in time conditional on $\boldsymbol{X}$ and $\boldsymbol{\Theta}$, but we can imagine a system that evolves according to a trending or periodic governing process not accounted for by S, and therefore for there to be some temporal correlation in e independent of $\boldsymbol{X}$ and $\boldsymbol{\Theta}$. We could condition e on more and different types of current information, for example, model states [e.g., *Wilkinson et al*., 2011], which might help further reduce the impacts of nonstationary model error, but nothing will ensure that statistical properties of model error discovered during calibration will persist in predictions.

Because of the dimensionality problem, it is impractical to nonparametrically estimate model error distributions (e.g., without autocorrelation assumptions), and model error may be nonstationary with respect to any distribution that we do estimate. MK attenuate these issues but do not alleviate them, and attenuation does not come from the fact that they do not evaluate f_e, but from the fact that f_e is semiempirical and conditional on $\boldsymbol{X}$. The ability to use an empirical f_e distribution represents a fundamental advantage over methods such as GLUE [*Beven and Binley*, 1992] that need an understanding of model error to derive $f_{\Theta|D}$; however, the problem is that we cannot estimate $f_{D|\Theta}$ if we do not know $f_{e|D}$ and we cannot estimate $f_{e|D}$ if we do not know $f_{\Theta|D}$—we really want their joint distribution.

MK argue that it is not strictly necessary to use $f_{D|\Theta}$ to estimate $f_{\Theta|D}$ [*Ebtehaj et al*., 2010; *Srikanthan et al*., 2009], which is true because fundamentally there are two ways to estimate conditional probability distributions: *generatively* and *discriminatively* [*Nearing et al*., 2013]. Generative methods estimate the joint distribution via Bayes' law: e.g., $f_{\Theta,D}=f_{D|\Theta}f_{\Theta}$, and discriminative methods employ a functional approximation of the desired conditional: e.g., $f_{\Theta|D}(\boldsymbol{\Theta}|\boldsymbol{D})=g_D(\boldsymbol{\Theta},\boldsymbol{D})$. Generative methods require likelihood functions and discriminative methods do not; however, both require estimating at least one function: either g_D or $f_{D|\Theta}$, and both are inherently subject to the same set of problems: subjectivity, nonstationarity, and dimensionality. For example, *Ebtehaj et al*. [2010] generated an ensemble of "optimal" parameter estimates by minimizing an objective function over bootstrapped samples of calibration data. If the objective function is a (negative-log) likelihood function, then this method provides a distribution over maximum-likelihood parameter sets and is generative. If the objective function is not a formal probability measure then the method is discriminative, results do not have a strict probabilistic interpretation, and the method suffers the same problem as GLUE—a subjective and informal likelihood. We still have not escaped the underlying problems.

A better way to address the problem of simultaneously estimating both $f_{\Theta|D}$ and $f_{e|D}$ is to estimate their joint distribution; in fact, we would like the joint distribution over e, $\boldsymbol{\Theta}$, and $\boldsymbol{X}$. *Harrison* [2007] and *Harrison et al*. [2012] estimated the joint distribution over e and $\boldsymbol{\Theta}$ while *Bulygina and Gupta* [2009, 2010, 2011] avoided the problem by using (1) to turn a deterministic model into a stochastic model and then conditioned the parameters of the stochastic model directly on data, thereby eliminating the need for a description of model error altogether. No study, to my knowledge, has attempted to build the full joint distribution.

In conclusion, MK lay out a compelling framework for estimating forecast uncertainty that directly attacks the central issues related to applying what is unequivocally the correct (Bayesian) method for combining uncertainty sources. Their approach of using an empirical $f_{e|D}$ conditional on current information ($\boldsymbol{X}$ and $\boldsymbol{\Theta}$) will certainly help attenuate effects of the major problems related to estimating epistemic uncertainty (subjectivity, nonstationarity, and dimensionality), and is consistent with other sophisticated contemporary efforts; however, avoiding evaluation of a likelihood function cannot offer fundamental solutions to these problems. MK's formulation does not really avoid the need to evaluate a likelihood function, and although this is possible by simply using discriminative methods to estimate conditional density functions, such a strategy does not inherently address the underlying issues.

References

Bellman, R. (2003), *Dynamic Programing*, Dover, Mineola, N. Y.

Beven, K., and A. Binley (1992), The future of distributed models—Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*, W00B13, doi:10.1029/2007WR006749.

Bulygina, N., and H. Gupta (2010), How Bayesian data assimilation can be used to estimate the mathematical structure of a model, *Stochastic Environ. Res. Risk Assess.*, *24*(6), 925–937, doi:10.1007/s00477-010-0387-y.

Bulygina, N., and H. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resour. Res.*, *47*, W05514, doi:10.1029/2010WR009614.

Cox, R. T. (1946), Probability, frequency and reasonable expectation, *Am. J. Phys.*, *14*, 1–13, doi:10.1119/1.1990764.

Ebtehaj, M., H. Moradkhani, and H. V. Gupta (2010), Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling, *Water Resour. Res.*, *46*, W07515, doi:10.1029/2009WR007981.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.

Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Philos. Trans. R. Soc. London A*, *222*, 309–368, doi:10.1098/rsta.1922.0009.

Harrison, K. W. (2007), Test application of Bayesian programming: Adaptive water quality management under uncertainty, *Adv. Water Resour.*, *30*(3), 606–622, doi:10.1016/j.advwatres.2006.03.011.

Harrison, K. W., S. V. Kumar, C. D. Peters-Lidard, and J. A. Santanello (2012), Quantifying the change in soil moisture modeling uncertainty from remote sensing observations using Bayesian inference techniques, *Water Resour. Res.*, *48*, W11514, doi:10.1029/2012WR012337.

Howson, C., and P. Urbach (1989), *Scientific Reasoning: The Bayesian Approach*, Open Court Publ., Chicago, Ill.

Kolmogorov, A. N. (1956), *Foundations of the Theory of Probability*, Chelsea Publishing Co., N. Y.

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

Montanari, A., and D. Koutsoyiannis (2012), A blueprint for process-based modeling of uncertain hydrological systems, *Water Resour. Res.*, *48*, W09555, doi:10.1029/2011WR011412.

Nearing, G. S., W. T. Crow, K. R. Thorp, M. S. Moran, R. H. Reichle, and H. V. Gupta (2012), Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment, *Water Resour. Res.*, *48*, W05525, doi:10.1029/2011WR011420.

Nearing, G. S., H. V. Gupta, and W. T. Crow (2013), Information loss in approximately bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity, *J. Hydrol.*, *507*, 163–173, doi:10.1016/j.jhydrol.2013.10.029.

Rathmanner, S., and M. Hutter (2011), A philosophical treatise of universal induction, *Entropy*, *13*(6), 1076–1136, doi:10.3390/e13061076.

Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, *44*, W03423, doi:10.1029/2007WR006357.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, *136*(12), 4629–4640, doi:10.1175/2008MWR2529.1.

Srikanthan, R., G. A. Kuczera, and M. A. Thyer (2009), 'Calibrate it twice': A simple resampling method for incorporating parameter uncertainty in stochastic data generation, paper presented at H2009: The 32nd Hydrology and Water Resources Symposium, Eng. Aust., Newcastle, N. S. W., 30 Nov. to 3 Dec.

Van Horn, K. S. (2003), Constructing a logic of plausible inference: A guide to cox's theorem, *Int. J. Approximate Reasoning*, *34*(1), 3–24, doi:10.1016/S0888-613X(03)00051-3.

van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: An extremely efficient particle filter, *Q. J. R. Meteorol. Soc.*, *136*(653), 1991–1999, doi:10.1002/qj.699.

Vrugt, J. A., and B. A. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci.*, *104*(3), 708–711, doi:10.1073/pnas.0610471104.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Vrugt, J. A., H. V. Gupta, and B. O. Nuallain (2006), Real-time data assimilation for operational ensemble streamflow forecasting, *J. Hydrometeorol.*, *7*(3), 548–565, doi:10.1175/JHM504.1.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environ. Res. Risk Assess.*, *19*(6), 378–387, doi:10.1007/s00477-005-0006-5.

Wikle, C. K., and L. M. Berliner (2007), A Bayesian tutorial for data assimilation, *Physica D*, *230*(1-2), 1–16, doi:10.1016/j.physd.2006.1009.1017.

Wilkinson, R. D., M. Vrettas, D. Cornford, and J. E. Oakley (2011), Quantifying simulator discrepancy in discrete-time dynamical simulators, *J. Agric. Biol. Environ. Stat.*, *16*(4), 554–570, doi:10.1007/s13253-011-0077-3.

Zupanski, M. (2005), Maximum likelihood ensemble filter: Theoretical aspects, *Mon. Weather Rev.*, *133*(6), 1710–1726, doi:10.1175/mwr2946.1711.