IMPEACTS OF OBSERVATIONAL UNCERTAINTY ON ANALYSIS AND MODELLING OF HYDROLOGICAL PROCESSES

An epistemically uncertain walk through the rather fuzzy subject of observation and model uncertainties

Keith Beven

Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence
Keith Beven, Lancaster Environment Centre, Lancaster University, Lancaster, UK.
Email: k.beven@lancaster.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/R004722/1

1 | SETTING OFF ON THE WALK

Way back in the 1970s, when my wanderings in this area began, uncertainty was simple. It was statistics and the probability calculus (and some will argue, i.e., still all we require but we will come back to that). It was manifest in various forms, from the use of standard errors of estimation in regression in the analysis of data (though these were often not cited when the regressions were actually used) to sampling requirements to achieve a required level of confidence in the mean of an observable (though these requirements were often not met in actually taking the samples). These were still the early days of digital computing applications in hydrology, with only limited computer power available, and one advantage of making simple statistical assumptions was that in some cases it was possible to derive analytical or semi-analytical solutions. At that time there was little consideration of observational uncertainties in hydrological data — papers concerned with uncertainty were mostly about sampling statistics (e.g., Hills & Reynolds, 1969 for near-surface soil moisture; and Huff, 1970, and Larson & Peck, 1974, for rainfalls).

The idea of uncertainty then was actually more often referred to in relation to potential future hydrological time series as stochastic variables (e.g., for optimization of water resource management, reservoir design and flood frequency analysis). How to generate stochastic time series using Monte Carlo methods was an important area of development as digital computers became more widely available (e.g., amongst many others, Colston & Wiggert, 1970; Cowpertwait & O’Connell, 1992; Harms & Campbell, 1967; Weiss, 1977).

Another interesting use of Monte Carlo simulation that started in the late 1970s was the generation of spatial patterns of parameters in distributed models, initiated by Al Freeze (1975, 1980) and Smith and Hebbert (1983). This was a response to the field observations of heterogeneity in soil moisture characteristics instigated by Don Nielsen and Jim Biggar (e.g., Nielsen et al., 1973; see also, e.g., Sharma et al., 1980; Russo & Bresler, 1980) and was also made possible by the availability of more computer power (Al Freeze was also working at the IBM Thomas J Watson research centre so had access to the biggest ‘mainframe’ computers around). Sample numbers tended to be low and the first studies of Monte Carlo realizations of hydraulic conductivities were limited to a single dimension (Freeze, 1975) before it was realized that this artificially exaggerated the effect of low conductivity regions because they could not be by-passed. One of my first contributions to this literature was a comment on the paper by Freeze (1980). In that comment I suggested that the question was not really one of evaluating the potential variability of stochastic realizations, but rather finding the single realization that best represented the real system (Beven, 1981). That continues to be the challenge in hydrology and, in one sense, that is what I (and many others) have been trying to do with various models ever since. However, it has been necessary to recognize that the dimensionality of the problem is such that finding that realization is necessarily constrained by the uncertainty in both observations and model structures so that, at best, it will only be possible to find a sample of plausible representations of that reality (albeit that these might be difficult to identify by prior assumption alone). This is what underlies the concept of
equifinality of model realizations in the face of the different sources of uncertainty (Beven, 2006, 2019c).

2 | EPISTEMIC OBSERVATIONAL UNCERTAINTIES

I also realized at an early age that our situation as hydrologists is actually worse than that because what we are dealing with is a lack of knowledge (or what is now often referred to as epistemic uncertainty) about the real system as a result of our limited observational capability. Epistemic uncertainty then allows the possibility of many different interpretations of the meaning of the available observations in describing reality. An obvious example is the estimation of rainfall over a catchment area. When based on raingauges, this means some recognition of the potential uncertainties at a measurement site and also requires some interpolation procedure that may involve other variables such as functions of elevation. When based on radar reflectivities or the processing of satellite data, the estimation becomes even more complex (and time-variable) but, again, the resulting ‘observations’ are often quoted without any consideration of the uncertainty. To allow for uncertainty, both types of estimation will require some assumptions, or an observational model. The same will be true for discharge measurements, which are commonly based on a transformation of stage into discharge and, as such, might also be subject to epistemic uncertainties in the form of nonstationarities and extrapolations (e.g., Beven & Westerberg, 2011; Coxon et al., 2014; Di Baldassarre et al., 2016; Hollaway et al., 2018a; Juston et al., 2014; Kiang et al., 2018; McMillan et al., 2010, 2012, 2018; Westerberg et al., 2011). Similar issues apply to observations of water quality variables and their interpolation in time and space (e.g., Harmel et al., 2006, 2009; Hollaway et al., 2018a; McMillan et al., 2012; Montgomery & Sanders, 1986).

These epistemic observational uncertainties will consequently and necessarily have an impact in all the steps of the modelling process: in defining a perceptual model of the system; in the approximation of the perceptual model as equations; in the approximate numerical solution of those equations; in the evaluation of (ensembles of) model output variables against observations; and in using the model to make predictions with new boundary conditions (e.g., Beven, 2012a, 2012b). The common default assumption in past hydrological modelling of taking observations as true representations of real variables is not the best interpretation in that it may result in biased inference and prediction for practical applications (e.g., Beven, 2019a; Beven & Smith, 2015; Coxon et al., 2014; McMillan et al., 2010, 2018). Of course, if we had better knowledge, we might also be able to characterize and constrain the uncertainties in a better way. This includes defining the catchment and its river network as the object of study in the first place when there is a lack of knowledge about subsurface flow processes (e.g., Condon et al., 2020; Durighetto et al., 2020). There has also been the response that uncertainty is not actually a useful concept and that we should rather aim to see how much information can be extracted from the observational data (e.g., Nearing et al., 2016, 2020). This is where models of different types become involved, including the perceptual models hydrological processes models of uncertainties in the modelling process, and how to evaluate information in different contexts (Beven & Chappell, 2020; McMillan et al., 2018; Westerberg et al., 2017).

In contrast, it has long been recognized in statistical approaches to uncertainty and state space modelling that an observation cannot be taken as a true value of a variable. An observational model is really required, even if in hydrology many studies simply lump that uncertainty into a model residual. An observational model can be as simple as an additive error term, commonly assumed to be drawn from a normal distribution (as much for mathematical convenience as for any real understanding of the error). For hydrological observations we might need something more complex. Water level observations can be relatively precise (though not always in post-flood surveys), but when used to estimate discharges a regression equation is often used which will have its own associated (and potentially nonlinear) error model. This is often neglected when reporting discharge values and may be large when extrapolating beyond the range of any measured stage-discharge pairs. Discharge observations are, in a sense, virtual variables (e.g., Beven et al., 2012; Kiang et al., 2018; Coxon et al., 2015) and allowing for uncertainties can have an impact on decision making (McMillan et al., 2017). More recently hydraulic methods are increasingly used to estimate stage discharge relationships (e.g., Lang et al., 2010; Mansanarez et al., 2019), but these will be also subject to epistemic uncertainties resulting from changes in the 3D channel geometry and effective roughness coefficients (including any subsurface vegetation). This can then have an impact on consequent hydrological analyses, such as flood frequency estimation (e.g., Di Baldassarre et al., 2012; Fuentes-Andino et al., 2017).

3 | OBSERVATIONAL UNCERTAINTY, OPTIMIZATION, EQUIFINALITY AND GLUE

My interest in all things uncertain in hydrology stems from when I first starting doing Monte Carlo experiments with Topmodel in around 1980 when I was working at the University of Virginia and had access to the University's CDC 6600 mainframe computer. The output from those runs (or at least a subset of it) was printed on lineprinter paper but could also be stored as files on magnetic tapes. The tapes could be reloaded for further analysis, but there were still logistical difficulties in analysing the outputs from many thousands of model runs. It was a very slow process! One of the simplest things to do was simply to look at a global performance measure for each run, such as the Nash-Sutcliffe Efficiency that had been around for about a decade and which had been used widely in model optimization (Nash & Sutcliffe, 1970), though even in the first Topmodel paper (Beven & Kirkby, 1979) we had shown that optimization could lead to the model getting a good result using the ‘wrong’ process representation (see also Beven et al., 2020).

Those early Monte Carlo runs taught me that there was no clear optimal parameter set or model realization in practice in searching for a representation of the real system. Plots of performance measure
against parameters (later called dotty plots when used in the generalized likelihood uncertainty estimation [GLUE] methodology) often showed many equivalent models at or close to some upper limit of performance. The first dotty plots and the first variograms I plotted were actually made up of dots on lineprinter output! It was also clear that different performance measures resulted in different shapes for the response surface, but still often with some upper limit of performance (and significant model residuals, see also Andrássian et al., 2007). In similar studies elsewhere, notably in the group of Sor-oosh Sorooshian at the University of Arizona, this led to research on developing better ways to find the optimum, and the use of the Pareto front as a way of trading off the performance against different measures (e.g., Yapo et al., 1998). After the first publications appeared, GLUE itself was sometimes misunderstood as a type of Monte Carlo optimization method.

It took more than a decade to build up the confidence to publish the first GLUE paper (Beven & Binley, 1992, 2014). In part this was because, in talking to statisticians and statistical hydrologists, they could not see that there was a problem. There was well-established theory for estimating the uncertainty in model outputs by calculating, either analytically or numerically, the gradient of the response surface around the optimum or maximum likelihood solution. If that surface looked flat, it just indicated a lack of sensitivity to that particular parameter dimension. GLUE outcomes were criticized for not using a formal statistical likelihood or, as Bayesian statistics started to take hold, for not making stronger prior assumptions about parameter distributions and their covariance. At this time, the residual series from a model run was often referred to as ‘observation error’ because the analysis was conditional on an (often implicit) assumption that the model was correct (despite the George Box, 1976, aphorism that all analysis was conditional on an (often implicit) assumption that the model was correct). In response (Beven et al., 2008) we noted that formal likelihoods (including the correctness of the model) were known to be correct. In response (Beven et al., 2008) we noted that their examples were ideal examples, but real applications were non-ideal in that simple statistical assumptions about model residuals were difficult to justify. Even if we could assume, for example, that the observation errors on the inputs used to force a rainfall-runoff model were unbiased and Gaussian distributed (albeit that might be unlikely) any consequent model residuals would be necessarily complex in structure having been processed through a nonlinear model. This is one reason why trying to invert the process to identify error characteristics of rainfalls (as in the BATEA approach of Kuczera et al., 2006; Thyer et al., 2009; McMillan et al., 2011) results in very strong interactions between model parameters and event multipliers that will be conditioned by any model structural error.

Attempts to use a formal Bayes statistical framework in GLUE in both rainfall runoff modelling and hydraulic modelling required strong (by which I mean rather simple minded) assumptions about the nature of the residuals (treated as ‘observation error’ in the sense above) needed to define a formal likelihood (Romanowicz et al., 1994, 1996). Typical assumptions were that the residuals were unbiased and Gaussian distributed (sometimes with a constant lag 1 correlation). In applying likelihood theory in this way, more complicated structure in the residual series (and the way in which it might vary with model realization) was often ignored. In hydrological modelling, the nature of the residuals will often be structured in ways that differ between rising limbs and recessions, and timing errors can give strong autocorrelation in the residuals that might change between events and parameters sets. We would now interpret this as suggesting that the epistemic errors are dominating any simple random or aleatory variability. This was certainly not just a result of a model of observation uncertainty alone. Thus, the strong assumptions of a simple stationary likelihood function are often not met (though a common comment from statisticians has been that this problem was simply a matter of finding a likelihood function with better characteristics that could then be used in Bayesian theory).

One additional consequence of assuming a simple textbook likelihood function is that when used with large numbers of data points (as is often the case in hydrograph simulation for example), this produces highly stretched response surfaces (relative differences of many orders of magnitude) that did not seem (to me at least) to reflect real differences between model performances given expectations about the observation errors on both inputs and evaluation data. This induced me to use more informal likelihood measures (e.g., as proportional to the NSE or sum of absolute errors) that resulted in less stretching. I also started to explore fuzzy measures in a possibilistic framework as an alternative (see e.g., Franks et al., 1998; Freer et al., 2004; Hankin & Beven, 1998; Pappenberger et al., 2007). GLUE is generalized in allowing such choices, albeit that such choices will involve different logical and philosophical frameworks. Nearing et al. (2016) point out that not all such choices have a firm axiomatic foundation but my aim has always been to reflect common sense in model evaluation, and some of those results suggest that there are formal choices, such a formal statistical likelihoods combined using Bayes multiplication that might have a firm mathematical foundation but which led to overconfidence in model parameter estimates relative to the estimated variance of the residuals (see also more recently Beven, 2016; Beven & Smith, 2015).

However, the use of informal likelihood measures left applications of GLUE open to criticism for not applying statistical theory properly (e.g., Mantovan & Todini, 2006; Stedinger et al., 2008), including demonstrations on hypothetical examples where the assumptions of the formal likelihoods (including the correctness of the model) were known to be correct. In response (Beven et al., 2008) we noted that their examples were ideal examples, but real applications were non-ideal in that simple statistical assumptions about model residuals were difficult to justify. Even if we could assume, for example, that the observation errors on the inputs used to force a rainfall-runoff model were unbiased and Gaussian distributed (albeit that might be unlikely) any consequent model residuals would be necessarily complex in structure having been processed through a nonlinear model. This is one reason why trying to invert the process to identify error characteristics of rainfalls (as in the BATEA approach of Kuczera et al., 2006; Thyer et al., 2009; McMillan et al., 2011) results in very strong interactions between model parameters and event multipliers that will be conditioned by any model structural error.

That is not to say, of course, that formal likelihoods cannot give useful estimates of predictive uncertainty. The estimate of the error variance can always expand to be consistent with the magnitude of the model residuals found. There is, however, something deeply unsatisfying about using such strong assumptions about the residuals when it is clear that different model structures (and even parameter sets) might require different assumptions, and that the error structure can be changing in non-stationary ways, even over a single hydrograph. One response to that is to model that changing nature of the residual structure and make the nonstationary stationary (Koutsoyiannis & Montanari, 2015). In general, this will require more parameters of the residual model to be estimated, giving more degrees of freedom in allowing for structural deficiencies of the hydrological model, which is actually what we want to evaluate. But
because of the way in which the residual model can absorb the predictive limitations of a model, this statistical estimation process does not actually allow for model rejection, only for low likelihoods (Beven, 2016; Beven & Lane, 2019).

4 | EPISTEMIC UNCERTAINTIES AND INFEERENCE ABOUT PROCESSES

Is there any value in persisting with this epistemic/aleatory dichotomy? Clearly the boundary between them is not clear cut but is rather the result of the choices of the analyst. However, I think the value lies in focusing attention on the fact that some sources of uncertainty might be difficult to treat as simple probability distributions. Both ‘observed’ discharges and constructions of ‘observed’ rainfall over a catchment area are good hydrological examples of this. The probabilities might be really rather difficult to estimate, or might be incomplete, or the inferences from applying simple statistical theory might conflict with common sense. There is, of course, an argument that if the latter is the case then clearly the analyst has not made the correct assumptions from which to draw acceptable inferences; the problem has not been posed correctly. There are, however, so many examples of poor practice in the application of probabilities to hydrological modelling problems it suggests that overly simplistic application of probability theory can be delusional. I have, therefore, argued that we need some alternative mechanism for thinking more deeply about how to handle some of the significant knowledge uncertainties in hydrological inference, including those associated with our observations.

One of the issues that then arises is the role of epistemic uncertainties in inference about processes. As discussed elsewhere, we can qualitatively appreciate the complexities of hydrological systems through the use of perceptual models (Beven & Chappell, 2020). We can also analyse hydrological data in different ways to identify the dominant modes of response of hydrological systems. Much more difficult is infer how the underlying process complexity informs the dominant modes of response at the catchment scale (except for the obvious in terms of more or less baseflow dominated discharges, and even then there is ambiguity about what constitutes baseflow and the processes that support such longer time scale responses. Clearly, observational uncertainties and particularly epistemic observational uncertainties will make that process of inference, and the testing of process hypotheses, more difficult. Defining adequate hypothesis tests will then require some really hard thinking (see, e.g., Beven, 2019c; Beven & Lane, 2019).

5 | CAN OBSERVATIONS BE DISINFORMATIVE IN GETTING THE RIGHT RESULT FOR THE RIGHT REASONS?

A good example in this respect is the recognition that some observational data might be disinformative in the model evaluation process because of epistemic uncertainties in the observations irrespective of what model is being considered. In particular, while many process-based conceptual hydrological models are constrained to maintain mass balance, the data used to drive the model and evaluate the results might not be. This was first suggested by Beven et al. (2011) and later developed by Beven and Smith (2015) and Beven (2019a) based on event runoff coefficients which, surprisingly often, were sometimes greater than one (in some events much greater than 1), and sometimes unexpectedly low. This is not actually that surprising when we consider how the available observations are turned into model inputs while ignoring potential epistemic uncertainties in most applications. But no model that strictly maintains mass balance will predict a runoff coefficient greater than 1 or, under low flow conditions, build up apparent deficits greater than those specified as cumulative potential evapotranspiration. It is possible to incorporate some specific parameterization to modify rainfall or evapotranspiration inputs (both were allowed as multipliers in the Stanford Watershed Model back in 1964), or allow for subsurface exchange or abstractions to adjacent catchment areas, but the event-to-event variation in runoff coefficients is not necessarily explained in such terms; it is more epistemically uncertain. Thus, to use such an event in model calibration will lead to bias; to use it in evaluation or hypothesis testing might lead to incorrect inference.

More interestingly, when we come to predict the next event, for which we want to predict the output given some estimate of an input without the benefit of any observed output, we will not know whether the observations for that event will prove to be informative or disinformative, or how the antecedent conditions have been affected by observational errors in previous events (see the dual uncertainty bands in Beven & Smith, 2015). Beven (2019a) has recently suggested a method for using the empirical distribution of runoff coefficients to define limits of acceptability for model evaluation within GLUE, including those greater than one. While this is applicable only to catchments where events can be disaggregated by extrapolating recession curves, it is an example of thinking about the epistemic nature of sources of uncertainty in the modelling process and a response to the difficulties of treating epistemic input uncertainties in some explicit way. The approach does allow the limits of acceptability to be defined based on the available observations alone, without reference to any model runs.

This also means that it allows the possibility that none of the model runs will prove to be acceptable in this sense. This is not always the case in approaches based on model residuals; the variances in a statistical model of the residuals can always expand to allow greater predictive uncertainty without model rejection. But there are advantages in model invalidation; it means we have to do better in some way, either in checking the observational data for disinformative periods or in the representations of processes (Beven et al., 2018; Beven & Lane, 2019). It has been suggested, for example, that the fixed channel network assumption of many model formulations will not be able to simulate either the hydrographs or tracer observations in small catchments with networks that expand and contract within and between events, or where preferential flow pathways are
important in recharging the saturated zone (e.g., Beven, 2018). In both these cases, the potential for such processes has been recognized for a long time; the problem is that the information has not existed to allow some general parameterisations to be easily or satisfactorily incorporated into models.

6 | LOOKING TO THE FUTURE

Looking at the modelling problem from a rejectionist or invalidation viewpoint also points towards focusing on critical observations that might help distinguish between model formulations or parameterisations. This has been suggested for a long time (e.g., Beven, 2001) and is intrinsic in the discussions about having more interaction between the observational and computational hydrological communities (Beven & Chappell, 2020; Burt & McDonnell, 2015; Seibert & McDonnell, 2002). Progress in this direction has been slow, however: in part perhaps because of a general reluctance amongst modellers to declare their models invalidated, in part because of a lack of those critical observational techniques at scales where they can easily be compared with model predicted variables. The latter comparison is often fraught with commensurability uncertainties that are epistemic in that they inherently involve some lack of knowledge. Perhaps, however, given enough observations it might be possible to again define some limits of acceptability without resorting to the strong statistical assumptions of, for example, kriging interpolation that lack any link to how small-scale heterogeneities interact nonlinearly in producing responses at larger scales. The use of hydrological signatures, for example, might help in this respect in that as summary variables there will be some effect of integrating over uncertainties in the observations (e.g., Coxon et al., 2014; McMillan, 2020). This would, of course, work best if those uncertainties were aleatory rather than epistemic with nonstationary bias in nature.

There is an analogy here with Approximate Bayesian Computation (ABC), developed later in the field of genetics for high dimensional fitting problems where it was particularly difficult to define a likelihood function (e.g., Pritchard et al., 1999) but with roots in a more generalized statistical theory (Diggle & Gratton, 1984). ABC works by searching for models that give results within a certain tolerance criterion in matching the available observations. It can be shown (again for hypothetical ideal cases) that this can give a good approximation to formal Bayesian likelihood theory, at least when small tolerances are achievable. It has been used much more widely, however, with a variety of tolerance measures. An adaptive tolerance search strategy is often used, homing in on model parameter sets that give good fits as the tolerance gets smaller (as in, e.g., in DREAM_LoA algorithm of Vrugt & Beven, 2018). There will then be interaction between tolerance and model performance, as opposed to setting limits of acceptability prior to making any model runs based on what is known about the data. This is an important difference between ABC and GLUE limits of acceptability approach since it means that ABC applications will not allow the tolerance to be decreased to a point that results in model invalidation. I have argued that model invalidation provides a useful way of doing science in the world of the inexact sciences (Beven, 2019a). As such, it is a form of response to the criticism that focusing on uncertainty estimation leads to undermining the science (see Beven, 2006, 2008; and others), even if in making predictions with an ensemble of models we might still be underestimating the epistemic uncertainties (Andréassian et al., 2007).

This limits of acceptability approach is also analogous to a form of fuzzy reasoning for which the foundations have been provided by authors such as Lotfi Zadeh, Joseph Halpern and George Klir. Within the limits of acceptability, fuzzy membership weights can be defined and combined in a variety of ways to reflect what is considered to be a better fit to the data. Prior membership functions can be defined to guide sampling model spaces in ways analogous to Bayesian prior probabilities. Of course, it is also possible to be thoughtful within a probabilistic framework, and in ways that are not based on model residuals (e.g., the comparison with data-based models used by Nearing et al., 2016, 2020; Ruddell et al., 2019).

As I noted in Beven (2019b), I am ending my research career with much more uncertainty than when I started. In retrospect this is not surprising; even a cursory examination of the type of hydrological data we use in modelling suggests that epistemic uncertainties are generic, even if we did not (naïvely) really think about this back in the 1970s, and even if such uncertainties are still commonly ignored in hydrological analyses and model calibrations (see the discussion in Pappenberger & Beven, 2006; there has been only limited change since). At least over the period of my career we have become somewhat less naïve about both observational and modelling uncertainties (helped by the increase in computer power that has allowed the exploration of some of the consequences). What is clearer now is that the thoughtless application of recipes (such as residual-based formal likelihood measures) can be misleading and lead to overconfidence in inference. Hydrologists in general need to be more thoughtful in their analyses and learn much more about their observational data when testing models as hypotheses about how catchments function. In this respect, some flexibility will be required as the information content of the observations will surely be different from application to application and for different prediction requirements in the same catchment.

Finally, I would suggest that we need to accept that we are dealing with an inexact science, for which a starting point is assessing the uncertainty associated with the observations (before we even start to worry about modelling issues). However, because of the epistemic nature of the observational uncertainties there can be no correct answer, only those consistent with some assumptions about how to approach the problem whether probabilistic, possibilistic or some framework based on the utility of the information in the observations provided for a particular purpose. If there is no agreement about how uncertainty should be handled then this is for good epistemic reasons (Beven, 2008, 2019a; Juston et al., 2013; Westerberg et al., 2017) – if we had more information about the limitations of hydrological observations then surely it would be easier to come to a consensus (see Beven, Asadullah, et al., 2020). What is important is to remember that, regardless of framework, uncertainty estimation is only a means to an
end: it is only a way of establishing the degree of confidence we might have in using the available observations to assess the decisions required for a particular purpose.

One advantage of this approach is that in recognizing that there is no single correct approach to uncertainty estimation, it encourages the discussion and recording of assumptions. This can both facilitate communication with users of the outputs and provide an audit trail for the critical evaluation of those assumptions (Beven et al., 2014; Beven, 2018; Kiang et al., 2018), including assumptions about observational uncertainties. It also does not preclude the choice of simple probability distributions and likelihood functions where they can be justified. This evaluation of the observational uncertainties is surely essential to anything that we might do in testing models as hypotheses (e.g., Beven, 2018). This includes the increasing availability of spatial datasets from remote sensing, cheap sensor networks, and citizen science. Each will be associated with epistemic observational uncertainties so that we will need to reflect on what those data actually mean, particularly relative to model variables that might not be commensurate, even if they have the same name. Beven and Lane (2019) have suggested an invalidation approach to testing models as hypotheses might while taking account of epistemic uncertainties. Fundamental to this is avoiding making the error of rejecting a good model just because of uncertainties in the observations. The papers in this special issue are surely an important step in exploring such issues.

There is, of course, value in observing with a view to improving the perceptual model of hydrological processes; to generate process hypotheses rather than testing models, using abductive rather than inductive reasoning (Baker, 2017; Beven & Chappell, 2020; Burt & McDonnell, 2015). In the past, this has generally been particularly fruitful when a new observational technique has become available, such as the change of paradigm arising from the availability of environmental isotope observations in the 1970s. What the hydrological community has not been so good at is generating or commissioning new observational techniques (Beven et al., 2020). To do so, of course, requires a demonstration of the value of the new data. Indeed, even in a semi-arid environment, it turned out to be not a lot.

There is no doubt that in hydrology we need to improve observational techniques, even for basic variables such as the precipitation inputs over a catchment (and particularly for snow and in hilly terrain). A degree of progress would certainly be possible by making existing technologies cheaper and more pervasive, but we should also aim to develop new methods, which means persuading funders that such methods would be useful, either scientifically in throwing more light on hydrological processes, or practically in improving model simulations and reducing uncertainties that feed into decision making. For that case to be convincing some quantification of the value of different observables will be necessary. In that quantification is necessarily model dependent, developing an accepted protocol might not be easy, but the final message from this brief history of walking with uncertainty is that if we do not start now then we will be waiting for another decade and probably longer.

ACKNOWLEDGEMENTS
I am grateful to all the guest editors of this special issue, Hilary McMillan, Ida Westerberg, Gemma Coxon and Anna Sikorska-Senoner, who carefully read and suggested significant improvements to earlier versions of this Commentary. This commentary has been produced within the Q-NFM project (NERC Grant NE/R004722/1) led by Dr. Nick Chappell at Lancaster University.

ORCID
Keith Beven ORCID: https://orcid.org/0000-0001-7465-3934

ENDNOTES
1 With an acknowledgement to Brian H. Kaye who wrote A random walk through fractal dimensions (Kaye, 1989).
2 We will name no names here, it was a long time ago and appreciation of the impact of epistemic uncertainties has changed a lot.
3 Again, it is not necessary to name names.
4 Interestingly this was developed at Lancaster University by Peter Diggle shortly before I took up a position there in 1985. The similarity between that theory and the GLUE methodology was not recognized when I gave a seminar in the Department of Statistics at Lancaster not long after – surely because of the nature of my presentation, having mislaid on the day all the overhead transparencies I had so carefully prepared in the pre-Powerpoint era!

REFERENCES
Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Oudin, L., & Perrin, C. (2007). What is really undermining hydrologic science today? Hydrological Processes, 21, 2819–2822.
Baker, V. R. (2017). Debates – Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty. Water Resources Research, 53(3), 1770–1778.
approach. In V. Barnett & K. F. Turkman (Eds.), Statistics for the Environment II. Water related issues (pp. 297–317). Wiley.

Romanowicz, R., Beven, K. J., & Tawn, J. (1996). In M. G. Anderson, D. E. Walling, & P. D. Bates (Eds.), Bayesian calibration of flood inundation models (pp. 333–360). Floodplain Processes.

Ruddell, B. L., Drewry, D. T., & Nearing, G. S. (2019). Information theory for model diagnostics: Structural error is indicated by trade-off between functional and predictive performance. Water Resources Research, 55(8), 6534–6554.

Russo, D., & Bresler, E. (1980). Scaling soil hydraulic properties of a heterogeneous field. Soil Science Society of America Journal, 44(4), 681–684.

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multi-criteria model calibration. Water Resources Research, 38(11), 23.1–23.14.

Sharma, M. L., Gander, G. A., & Hunt, C. G. (1980). Spatial variability of infiltration in a watershed. Journal of Hydrology, 45(1–2), 101–122.

Smith, R. E., & Hebbert, R. H. B. (1983). Mathematical simulation of interdependent surface and subsurface hydrologic processes. Water Resources Research, 19(4), 987–1001.

Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research, 44(12), W08B06. https://doi.org/10.1029/2008WR006822.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., & Srivastava, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research, 45(12), W00B14. https://doi.org/10.1029/2008WR006825.

Vrugt, J. A., & Beven, K. J. (2018). Embracing equifinality with efficiency: Limits of acceptability sampling using the DREAM(LOA) algorithm. Journal of Hydrology, 559, 954–971.

Weiss, G. (1977). Shot noise models for the generation of synthetic streamflow data. Water Resources Research, 13(1), 101–108.

Westerberg, I., Guerrero, J.-L., Selbert, J., Beven, K. J., & Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. Hydrological Processes, 25, 603–613. https://doi.org/10.1002/hyp.7848.

Westerberg, I. K., Di Baldassarre, G., Beven, K. J., Coxon, G., & Krueger, T. (2017). Perceptual models of uncertainty for socio-hydrological systems: A flood risk change example. Hydrological Sciences Journal, 62(11), 1705–1713.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. Journal of Hydrology, 204(1–4), 83–97.

How to cite this article: Beven K. An epistemically uncertain walk through the rather fuzzy subject of observation and model uncertainties. Hydrological Processes. 2021;35:e14012. https://doi.org/10.1002/hyp.14012