Hydrological Processes

COMMENTARY OPEN ACCESS

Deep Learning, Epistemic Uncertainties and the Inexact Sciences

Keith Beven

Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence: Keith Beven (k.beven@lancaster.ac.uk)

Received: 20 February 2026 | **Accepted:** 6 March 2026

Keywords: disinformation | epistemic uncertainty | machine learning

ABSTRACT

This commentary considers the recent explosion of models based on Deep Learning/Machine Learning methods in hydrological journals and how their application might better contribute to hydrological understanding in the face of epistemic uncertainties in an inexact science.

It seems that every contents list of every hydrological journal is now dominated by papers that make use of methods of machine learning or deep learning (DL), encouraged, it seems, by the way in which DL models have demonstrated improved predictive ability relative to more conceptual modelling approaches and the ready availability of software (and perhaps just because it is fashionable and DL software is readily available). Improved predictive ability by some measure is, of course, a very important result. It shows that we have more to learn about the representation of hydrological processes—or also about the uncertainties and errors in the hydrological data that are used to train the models.

Because, as yet, there is no real explanation of what is leading to this improved predictive ability (although it has been shown that adding hydrological constraints to DL models might add little to predictive power—see, for example, Álvaraz Chaves et al. 2026). It is also still the case that, in applications of DL to large sample catchment data sets, there is still a significant number of catchments where the fit is poor and the model cannot be considered as fit-for-purpose (as discussed previously in Beven 2023, see also Figure 5 in Álvarez Chaves et al. [2026] for the CAMELS-UK dataset as a recent example).

This does perhaps suggest that, at least in some catchments, DL methods can to some extent compensate for limitations in understanding of processes and the inexactitudes of hydrological data in ways that we do not yet properly understand. For hydrology is, indeed, one of the inexact sciences (Beven 2019a). It deals with processes that mostly cannot be properly observed at the scales of predictive interest. It makes observations that do not adequately represent the fluxes of interest (including actual inputs to catchment areas and evapotranspiration losses over heterogeneous catchment areas). One of the most important fluxes, stream discharge, is almost always not measured directly, but inferred from measurements of water levels with all the potential uncertainties that can involve, particularly in channels with mobile beds and during overbank flows (Beven and Westerberg 2011; Beven et al. 2012; McMillan et al. 2012, 2018). Other subsurface discharges from a catchment area may not be measured at all. Thus, any model representation of hydrological processes will necessarily also be inexact. These are, in fact, all sources of epistemic uncertainties in hydrology analysis and prediction. Epistemic uncertainties are those that are the result of a lack of knowledge, rather than uncertainties that can be represented as 'irreducible' aleatory statistical variability.

Let us restate some of the implications of accepting that epistemic uncertainties are an issue in hydrological analysis and prediction. The nature of epistemic errors means that there may be inconsistencies in data series that, in interaction with the epistemic error of model structural error, will affect any model calibrations and that may not be reduced by integration over a time series as might be expected by aleatory statistical errors. For example, nearly all modelling papers do not consider that some of the data available might be *disinformative* for the purpose of model evaluation (e.g., where events have apparent runoff coefficients greater than 1, see Beven and Smith 2015; Beven 2019a, 2024). As a result, we should not expect there to be any clear optimal model, but rather many different model formulations that might be acceptable in relation to the uncertainties in the data (the *equifinality thesis* of Beven 2006). We also expect that different epistemic errors in different sections of the available time series, or different objective functions and observables used in evaluation in interaction with the epistemic errors, will result in different, sometimes quite different, optimal models for a given catchment (and the same will apply for large sample catchment data sets).

It has been suggested to me recently that optimised DL models are a solution to the equifinality problem but I cannot see how this can be the case, particularly when their many, many parameters are optimised with respect to the Nash-Sutcliffe Efficiency (with its known deficiencies, e.g., Melsen et al. 2025), as is still the case in many recent studies. While such models might indeed produce higher values of efficiency, how can they be immune from the impacts of epistemic error on the optimisation? Is it only the case that, relative to conceptual models, they can identify some consistency in the data uncertainties that allows them to overfit to the errors? And, with so many parameters or weights to be identified, why should it not the case that there will be other sets of weights (or internal structures) that will be very nearly as good in fitting the training data as that chosen as the optimum. As far as I know, no studies have yet evaluated the possibility for equifinality in sets of weights or structures in DL models—but that does not mean that allowing for data uncertainties will not reveal similar equifinalities to more traditional models with many less parameters and consequently less flexibility.

It is perhaps worth re-emphasising that hydrology as an inexact science is revealed more in those catchments where even DL models fail to provide good predictions. This tail of poorly fitting catchments is surely the most interesting result hydrologically but is mostly totally ignored in the discussion of the results. It is the most interesting because, particularly in the inexact sciences, it is the case that we will learn more from considering failures than from successes. We do not learn much from accepting an optimal model as fit-for-purpose if that model is difficult to interpret or if there are some important aspects of the model predictions that do not reproduce what is required for the purpose at hand.

So, if we wish to progress the science, it is far more important to pose the question just why it is that there are catchments for which DL (and other conceptual hydrological models) fail to produce acceptable predictions? This is certainly primarily a matter of epistemic uncertainties in the data sets. In semi-arid and arid catchments, for example, that may simply be a matter of how a limited number of rain gauges cannot represent the pattern of inputs, even in experimental catchments. It may also be the case that a catchment scale model structure, with catchment averaged inputs, cannot reproduce the complex spatial patterns of antecedent conditions that, in interaction with the pattern of input rainfalls in space and time, will control the runoff generation processes. This is, again, another form of epistemic uncertainty. The important point here is that these failures need to be investigated as part of the learning process so as to understand how we might be able to do better in future.

This, of course, is not new. It has been stated in different ways by myself and others in the past (e.g., Beven 2020, 2021, 2023). The odd thing is how it continues to be overlooked in papers that do not take any account of the implications of epistemic uncertainties. There are recent papers that continue to present the results only of 'optimal' DL models. There are recent papers that continue to treat uncertainties only as aleatory variability. There are still modelling papers that do not mention any uncertainties associated with their predictions at all.

How then can we improve practice as the use of DL methods matures, and as a consequence improve our hydrological science? Two important improvements would be a much closer examination of hydrological data for modelling purposes, and of model results as fit-for-purpose (Beven 2019b; Beven and Lane 2022). Clearly, that has not been the practice in the application of DL models, which suggests that there is, in fact, more to learn about both data and models to understand how DL can compensate for deficiencies in data and adapt to different objective functions. There also needs to be improved process representations in hydrological models. In this respect there will surely be a role for DL models as we gain understanding of how best to use them (see the 'Period of Consolidation' suggested in Beven 2025a and the Great Hysteresis Challenge of Beven et al. 2026). That understanding will need to address the problem of how to disaggregate the epistemic uncertainties in the modelling process into those associated with structural uncertainties in the representation of processes and those associated with some consistent data uncertainties (noting that if the data uncertainties are purely inconsistent then DL will find no information content for improved prediction).

This does not matter, of course, if the aim of the modelling is only improved prediction for purposes such as adaptive flood forecasting or infilling of data series, when the means of improved prediction is really not important (even if residual uncertainties might still be important to decisions). It does matter if, on the other hand, we are interested in predictions of changed conditions, whether that be changes to the inputs or changes in the catchment itself, which might result in behaviours outside the range of the training data. Somewhat amusingly with the benefit of hindsight, there is an analogy here with the similar argument made many decades ago in favour of the use of distributed 'physically-based' models of hydrological processes (see, e.g., Beven and O'Connell, 1982). That argument was based on the idea that such distributed models would be a better representation of the processes in a catchment, and the idea that the parameters of the model, as 'physically-based' would be much easier to change to represent changing characteristics in a more realistic way. Neither of those ideas proved to be correct—the process representations were not sufficiently realistic and the effective values of the parameters too difficult to estimate with any certainty.

There is a resonance here with the recent work of Álvarez Chaves et al. (2026), who suggest that when DL is allowed to vary the parameters of conceptual models in predicting discharges, little of the predictive capability comes from the physical constraints imposed by the conceptual model structures. This really suggests that DL is using parameter variations to compensate for both model structural limitations and epistemic errors in the input and calibration data. We can also draw an analogy here with the earlier studies of Thyer et al. (2009) using Bayesian Total Error Analysis (BATEA) which used event rainfall multipliers to improve model performance. The result was a distribution of multipliers spanning a wide range (0.5 to > 3) that really appeared physically unrealistic, but which necessarily reflected the interactions between model structural limitations and epistemic errors in the input and calibration data. In this case, that interaction somewhat was more easily interpretable than the implicit nature of the compensation in the DL models, though it is clear from the parameter value time series plots in Álvaraz Chaves et al. (2026) that there is an interaction between the parameter values and the hydrological state of the catchment. That information is being exploited by the DL models.

We need to understand these compensating effects, not simply declare success as improved predictive performance on some (rather arbitrary) objective function, but to understand what is leading to how that information is being exploited. One issue is that epistemic errors in the input data can be compensated by an adjustment to either parameters or event rainfall multipliers. Underprediction can be compensated by a change in parameter value(s) or by increasing the rainfall multiplier. Overprediction can be compensated by a change in parameter value(s) or by decreasing the rainfall multiplier. This then means that timing is important because those changes will then affect how well a model will be able to produce the next event (and subsequent events). But a lot of that timing information will be lost when models are fitted on the basis of global objective functions such as KGE or NSE. Disaggregating and understanding such compensating effects over time is clearly a very challenging problem (and made worse if we allow for any equifinality of model representations), but it is surely not a problem that should be ignored.

So, what should be the priorities in trying to do so? I would suggest that the greatest priority is to take much more care about evaluating hydrological data and trying to identify epistemic uncertainties in hydrological data before any modelling exercise of whatever type. This is not standard practice. We take any estimates of catchment average or gridded rainfall inputs, estimates of snow water equivalents, or discharge observations generally supplied without any associated estimates of uncertainty as given and 'true'. This is perfectly understandable for many practical applications in hydrology where there is neither time nor money available to do more. But it should not be the case in research applications where the aim is greater understanding of the modelling process. In doing so, assessing the potential for epistemic errors in the data used for modelling is surely fundamental to any better understanding. A good place to start in this respect would be by trying to understand those catchments where the fit of DL models remains poor.

A second priority is perhaps to improve our model evaluations in terms of what is fit-for-purpose given the assessment of uncertainties in the data (see Beven and Lane 2022; Beven 2023). This might mean concentrating on particular aspects of model performance for the purpose at hand (e.g., Beven et al. 2022) or on the use of different types of observable in model evaluations (e.g., Fang et al. 2013; Ehlers et al. 2019). Where those observables are spatially distributed then that might require spatially distributed models (with all the additional parameter dimensions that implies) to make full use of the additional information and the implications for the representation of hydrological processes (e.g., Beven 2023, 2025b).

One feature that is normally included in any definition of epistemic uncertainty is that it might be reduced by obtaining more understanding or knowledge about that uncertainty. Should this be an identified aim for DL in this 'Period of Consolidation'? It might be a valuable direction for further research since it is clear that the challenge of learning about hydrological epistemic uncertainties, and their disaggregation, has been largely neglected. While this is understandable because of the difficulties involved, and too late for me now as long retired, it is surely fundamental to making real advances in progressing the science.

Funding

The author has nothing to report.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

Álvarez Chaves, M., E. Acuña Espinoza, U. Ehret, and A. Guthke. 2026. "When Physics Gets in the Way: An Entropy-Based Evaluation of Conceptual Constraints in Hybrid Hydrological Models." *Hydrology and Earth System Sciences* 30: 629–658. https://doi.org/10.5194/hess-30-629-2026.

Beven, K. J. 2006. "A Manifesto for the Equifinality Thesis." *Journal of Hydrology* 320: 18–36.

Beven, K. J. 2019a. "Towards a Methodology for Testing Models as Hypotheses in the Inexact Sciences." *Proceedings of the Royal Society A* 475, no. 2224: 20180862. https://doi.org/10.1098/rspa.2018.0862.

Beven, K. J. 2019b. "How to Make Advances in Hydrological Modelling." *Hydrology Research* 50, no. 6: 1481–1494. https://doi.org/10.2166/nh.2019.134.

Beven, K. J. 2020. "Deep Learning, Hydrological Processes and the Uniqueness of Place." *Hydrological Processes* 34, no. 16: 3608–3613. https://doi.org/10.1002/hyp.13805.

Beven, K. J. 2021. "An Epistemically Uncertain Walk Through the Rather Fuzzy Field of Observational and Model Uncertainties." *Hydrological Processes* 35, no. 1: e14012. https://doi.org/10.1002/hyp.14012.

Beven, K. J. 2023. "Benchmarking Hydrological Models for an Uncertain Future." *Hydrological Processes* 37, no. 5: e14882. https://doi.org/10.1002/hyp.14882.

Beven, K. J. 2024. "A Brief History of Information and Disinformation in Hydrological Data and the Impact on the Evaluation of Hydrological Models." *Hydrological Sciences Journal* 69, no. 5: 519–527. https://doi.org/10.1080/02626667.2024.2332616.

Beven, K. J. 2025a. "The Seven Ages of Hydrology." *Hydrological Processes* 39: e70233. https://doi.org/10.1002/hyp.70233.

Beven, K. J. 2025b. "On the Future of Hydroecological Models of Everywhere." *Environmental Modelling & Software* 188: 106431. https://doi.org/10.1016/j.envsoft.2025.106431.

Beven, K. J., W. Buytaert, and L. A. Smith. 2012. "On Virtual Observatories and Modeled Realities (Or Why Discharge Must Be Treated as a Virtual Variable)." *Hydrological Processes (HPToday)* 26, no. 12: 1905–1908. https://doi.org/10.1002/hyp.9261.

Beven, K. J., and S. Lane. 2022. "On (In)validating Environmental Models. 1. Principles for Formulating a Turing-Like Test for Determining When a Model Is Fit-For Purpose." *Hydrological Processes* 36, no. 10: e14704. https://doi.org/10.1002/hyp.14704.

Beven, K. J., D. Mindham, and N. A. Chappell. 2026. "The Great Hysteresis Challenge." *Hydrological Processes*. https://doi.org/10.1002/jyp.70438.

Beven, K. J., and P. E. O'Connell. 1982. "On the Role of Physically-Based Distributed Models in Hydrology." Institute of Hydrology Report, No. 81, Wallingford, UK.

Beven, K. J., T. Page, B. Hankin, et al. 2022. "Deciding on Fitness-for-Purpose—Of Models and of Natural Flood Management." *Hydrological Processes* 36, no. 11: e14752. https://doi.org/10.1002/hyp.14752.

Beven, K. J., and P. J. Smith. 2015. "Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models." *Journal of Hydrologic Engineering* 20, no. 1: A4014010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991.

Beven, K. J., and I. Westerberg. 2011. "On Red Herrings and Real Herrings: Disinformation and Information in Hydrological Inference." *Hydrological Processes* 25: 1676–1680. https://doi.org/10.1002/hyp.7963.

Ehlers, L. B., T. O. Sonnenborg, and J. C. Refsgaard. 2019. "Observational and Predictive Uncertainties for Multiple Variables in a Spatially Distributed Hydrological Model." *Hydrological Processes* 33, no. 5: 833–848.

Fang, X., J. W. Pomeroy, C. R. Ellis, M. K. MacDonald, C. M. DeBeer, and T. Brown. 2013. "Multi-Variable Evaluation of Hydrological Model Predictions for a Headwater Basin in the Canadian Rocky Mountains." *Hydrology and Earth System Sciences* 17: 1635–1659.

McMillan, H., T. Krueger, and J. Freer. 2012. "Benchmarking Observational Uncertainties for Hydrology: Rainfall, River Discharge and Water Quality." *Hydrological Processes* 26, no. 26: 4078–4111.

McMillan, H. K., I. K. Westerberg, and T. Krueger. 2018. "Hydrological Data Uncertainty and Its Implications." *Wiley Interdisciplinary Reviews: Water* 5, no. 6: e1319.

Melsen, L. A., A. Puy, P. J. Torfs, and A. Saltelli. 2025. "The Rise of the Nash-Sutcliffe Efficiency in Hydrology." *Hydrological Sciences Journal* 70, no. 8: 1248–1259.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan. 2009. "Critical Evaluation of Parameter Consistency and Predictive Uncertainty in Hydrological Modeling: A Case Study Using Bayesian Total Error Analysis." *Water Resources Research* 45, no. 12: 1–22.