Developing observational methods to drive future hydrological science: can we make a start as a community?

Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., Cloke, H., Dadson, S., Everard, N., Fowler, H. J., Freer, J., Hannah, D. M., Heppell, K., Holden, J., Lamb, R., Lewis, H., Morgan, G., Parry, L. and Wagener, T. (2020) Developing observational methods to drive future hydrological science: can we make a start as a community? Hydrological Processes, 34 (3). pp. 868-873. ISSN 0885-6087 doi: https://doi.org/10.1002/hyp.13622 Available at http://centaur.reading.ac.uk/87260/

It is advisable to refer to the publisher’s version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/hyp.13622

Publisher: Wiley InterScience
Developing observational methods to drive future hydrological science: Can we make a start as a community?

Keith Beven1 | Anita Asadullah2 | Paul Bates3 | Eleanor Blyth4 |
Nick Chappell1 | Stewart Child5 | Hannah Cloke6 | Simon Dadson4,7 |
Nick Everard2 | Hayley J. Fowler8 | Jim Freer3 | David M. Hannah9 |
Kate Heppell10 | Joseph Holden11 | Rob Lamb12 | Huw Lewis13 |
Gerald Morgan14 | Louise Parry15 | Thorsten Wagener16

1Lancaster Environment Centre, Lancaster University, Lancaster, UK
2UK Environment Agency, Bristol, UK
3School of Geography, Bristol University, Bristol, UK
4Centre for Ecology and Hydrology, Wallingford, Wallingford, UK
5Hydro-Logic Services (International) Ltd., Reading, Reading, UK
6Department of Geography, University of Reading, Reading, UK
7School of Geography and the Environment, University of Oxford, Oxford, UK
8School of Engineering, Newcastle University, Newcastle upon Tyne, UK
9School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, UK
10School of Geography, Queen Mary University of London, London, UK
11water@leeds, School of Geography, University of Leeds, Leeds, UK
12JBA Trust, Skipton, Skipton, UK
13MetOffice, Exeter, Exeter, UK
14Edenvale Young Associates Ltd., Bristol, Bristol, UK
15Arup, Bristol, Bristol, UK
16Department of Civil Engineering, Bristol University, Bristol, UK

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

Hydrology is still, and for good reasons, an inexact science (for a recent discussion, see Beven, 2019a), even if evolving hydrological understanding has provided a basis for improved water management for at least the last three millennia. The limitations of that understanding have, however, become much more apparent and important in the last century as the pressures of increasing populations, and the anthropogenic impacts on catchment forcing and responses, have intensified (see Abbott et al., 2019; Montanari et al., 2013; Sivapalan, Savenije, & Blöschl, 2012; Srinivasan et al., 2017; Wagener et al., 2010; Wilby, 2019). At the same time, the sophistication of hydrological analyses and models has been developing rapidly, often driven more by the availability of computational power and geographical data sets than any real increases in understanding of hydrological processes. This sophistication has created an illusion of real progress, but a case can be made that we are still rather muddling along, limited by the significant uncertainties in hydrological observations, knowledge of catchment characteristics, and related gaps in conceptual understanding, particularly of the subsurface. These knowledge gaps are illustrated by the fact that for many catchments, we cannot close the water balance without significant uncertainty (e.g., Beven, 2019a;
Schaller & Fan, 2009), uncertainty that is often neglected in evaluating models for practical applications. This lack of water balance closure can also result from a lack of information about the influence of water management on the water balance. We have seen improvements since the first crude U.K. water balance estimates of John Dalton (1791), but there remain important uncertainties in the estimates of every term in the water balance equation: precipitation inputs (especially snow); discharge, evapotranspiration and other outputs; and storages in the system.

The above issues are reflected in the discussions that have produced the 23 unsolved problems in hydrology (Blöschl et al., 2019) and the British Hydrological Society Working Group on the Future of Hydrological Science (to which all of the co-authors have contributed). The aim of these two initiatives has been to stimulate hydrological research by identifying future strategic priorities. Here, we will focus on those areas pertaining to improving the understanding and representation of hydrological processes. Many of the unsolved problems refer to the nature and controls of future hydrological change, which surely requires a fundamental understanding of present-day hydrological processes and also of the human impacts on those processes (e.g., Abbott et al., 2019).

It could be considered that our perceptual understanding of hydrological processes is actually quite good (see, e.g., the outline in Beven, 2012), though, as in all the sciences, we still expect that understanding to improve over time. Examples of that improvement include recent work on the connectivity on hillslopes (e.g., Bracken et al., 2013; Emanuel, Hazen, McGlynn, & Jencso, 2014; Jencso & McGlynn, 2011) and the isotope studies that reveal differences in soil water and vegetation storages (McDonnell, 2014; Sprenger, Llorens, Cayuela, Gallart, & Latron, 2019). The difficulty comes in translating perceptual understanding, often gained in local experimental situations, into practical quantitative analyses of flows, storages, and water quality variables across a range of useful and appropriate time and space scales for a given purpose (see, e.g., the discussions in Beven, 2006; Beven & Germann, 2013; Ward & Packman, 2019). Quantitative analyses will require a model (even if it is only the water balance equation), and it is clear that the quantitative representation of hydrological processes in models is lacking in rigour because of the difficulty of testing models as hypotheses when the observational data are uncertain, at an inappropriate scale, or too sparse (e.g., Beven, 2019b; Beven & Lane, 2019). That is one reason why we have so many hydrological models. Current observational data are not adequate to reject many of our models (though see Holloway et al., 2018, for an example of the rejection of the rather widely used SWAT model).

To do better hydrology, we really need data streams for water fluxes, water storages, and water quality and catchment properties that will provide better inputs for hydrological predictions and support better hypothesis testing in improving hydrological science. That means better observational methods for all of the terms in the water balance equation as well as the tracer and quality observations required for a better understanding of residence time and transit time distributions and storage exchanges (see, e.g., Harman, 2015, 2019; Rinaldo et al., 2015; Remondi, Kirchner, Burlando, & Fatichi, 2018). Our current perceptual model allows for preferential flows, hot spots, hot moments, and other complexities in both surface and subsurface responses to forcing; most hydrological models do not include these and those that do have not been adequately tested as hypotheses. Scale is important here, since we do not fully understand how these small space-scale and timescale processes might integrate up to larger scales. What is clear is that such localized processes of recharge and run-off generation can be significant in affecting larger scale responses (e.g., Hartmann, Gleeson, & Wada, & Wagener, 2017; Fan et al., 2019; Ward & Packman, 2019).

But if we need to improve observations of all the water balance components, where to begin? This can, at least in part, be tested using simulations to establish what type of observational improvements might be more worthwhile for different purposes, for which different levels of uncertainty might be admissible. Within a Bayesian statistical framework, optimizing observational improvements can be explored using a form of pre-posterior prior analysis, where simulations are used to test the value of assuming new uncertain observations, or new types of observations, to be available. Such a framework has been used before in hydrology, for example, to assess where to place an additional observation well in assessing a groundwater model (see, e.g., Ben-Zvi, Berkowitz, & Kesler, 1988; Freeze, James, Massmann, Sperling, & Smith, 1992; Kollat, Reed, & Maxwell, 2011). In the remote sensing field, Observing System Simulation Experiments are similarly used to provide synthetic data sets for testing the utility of proposed missions (e.g., Durand et al., 2008; Biancamaria et al., 2011; in the case of the Surface Water and Ocean Topography [SWOT] satellite, still to be launched). The answers might not necessarily be simple. Bashford, Beven, and Young (2002), for example, looked at this type of observation gap problem from a slightly different perspective. In many parts of the world, including parts of the United Kingdom, evapotranspiration, rather than discharge, is the dominant output term in the water balance. Using simulations at a 30-m pixel scale, they produced a 1-km² scale evapotranspiration flux, which they assumed to be observed by remote sensing with different degrees of error. Using that spatial information, they explored what complexity of process model might be supported if such sensor signals could be made available. The outcome turned out to be much simpler than the representation of evapotranspiration in most hydrological models. This implies that both flux observations with low uncertainty and other types of information (e.g., internal states) would be required to support rigorous hypothesis testing to differentiate between model structures that reflect the complexity of processes in the environment. There will, inevitably, be a strong interaction between the development of model theory and the observational support available. The task then is to try to ensure that the right sort of data are collected for the purposes at hand, whether that be testing model structures or testing applied hydrological predictions.

As an example, one interesting possibility would be the development of a method for observing discharge in arbitrary channel cross sections but with sufficient accuracy to be able to identify spatial differences across the channel network. This spatial mapping is possible using tracers (see, e.g., Huff, O'Neill, Emanuel, Elwood, & Newbold,
1982; Genereux, Hemond, & Mulholland, 1993; Kelleher et al., 2013), at least under the assumption that the tracer is conservative (and with the permission of the relevant regulatory agencies). Such data collection might produce potentially significant improvements in understanding of the role of geology and topography in controlling hydrological processes (as also suggested by the inter-comparisons of catchment isotope responses in Tetzlaff, Seibert, & Soulsby, 2009; Birkel et al., 2018). It might also allow much stronger testing of distributed model predictions in ways not previously possible, subject to having adequate observations of the input forcing. Despite the large and ongoing investment in rainfall radar methods and better rain gauges, a major limitation on model testing is knowing just what the inputs to a catchment area are within complex terrain (e.g., Beven, 2019a; McMillan, Krueger, & Freer, 2012; Yatheendradas et al., 2008). If we did have better methods for estimating catchment inputs and discharges, we would be able to make much more rigorous hypothesis tests given information about storages and residence transit times.

Simulations will only go so far in deciding what type of measurements should be prioritized. That is because what is produced by a simulation depends on the structural assumptions of the model that produced it, and we have a mismatch between the complexity of the perceptual model of the relevant processes and the relative simplicity of current model structures. This mismatch will particularly be the case with estimates of storage and residence times that are strongly dependent on the assumptions underlying any simulation. Therefore, it would be worth combining the pre-posterior prior approach with some direct monitoring where intensive effort is made to capture all elements of the water balance accurately. There are some existing examples of intensive water balance monitoring in experimental systems such as the artificial hillslopes in Biosphere II in Arizona (Gevaert, Teuling, Uijlenhoet, & Troch, 2014; Hopp et al., 2009; Scudeler et al., 2016) or Hydrohill in China (Gu et al., 2010) where considerable effort is made to capture fluxes and storages. Intensive measurement of fluxes and storage is, however, difficult at larger scales in more “natural” catchments, as the long history of research on experimental catchments in different countries and climatic regions testifies. The International Hydrological Decade (1964–1975) generated a large number of “experimental or representative basins” globally (Robinson & Whitehead, 1993). Some, such as those at Plynlimon in Wales, are still being monitored and provide a strong case for the continuation of routine monitoring in a time of changing hydrological responses. Recent initiatives such as the TERENO basins in Germany (Bogena, 2016; Bogena et al., 2018), the Heihe basin in China (Li et al., 2013), and the CZO basins in the United States have seen considerable investment. Nevertheless, because of the limitations of current measurement technologies, and the lack of control over boundary conditions, it is not clear that any such experimental basin has the information available to critically test perceptual understanding and model formulations.

From an experimental viewpoint, we can consider new data collection of fluxes, storages, and catchment properties as an exercise in constraining uncertainty (Beven et al., 2018). In developing what is known about an existing experimental catchment, or in collecting data from a new study catchment (with a particular purpose in mind), we need to determine what types of information would be most useful in constraining the uncertainties in the understanding and prediction of the catchment responses necessary for that purpose, whether that be testing models as hypotheses or some decision for water management. Such an assessment would include making the most of information we might be able to bring from studies elsewhere (e.g., Evaristo & McDonnell, 2017), as well as information gained from direct observations, remote sensing, intensive field campaigns, or other strategies. The issue has been addressed in the context of the prediction of flow in ungauged basins (e.g., Blöschl, Sivapalan, Savenije, Wagener, & Viglione, 2013) but not in terms of considering the requirements for new observational techniques that might serve to improve hydrological science.

The latter purpose implies a need for better observational technologies and network designs to support hypothesis testing in real catchments of interest that go beyond current monitoring capabilities. This technological mission is necessarily long term because it does not seem that significant improvements to existing methods are yet on the horizon. There have been some improvements in radar and microwave rainfall estimates (Diederich, Ryzhkov, Simmer, Zhang, & Trömel, 2015; Rico-Ramirez, Liguori, & Schellart, 2015); eddy correlation and remote sensing estimates of evapotranspiration (Franssen, Stöckli, Lehner, Rotenberg, & Seneviratne, 2010; Maes, Gentine, Verhoest, & Gonzalez Miralles, 2019); gravity anomaly estimates of storage (Güntner et al., 2017; Huang et al., 2019; Riechey et al., 2015), acoustic Doppler measurements of discharge (Farina, Alvisi, & Franchini, 2017; Moore, Jamieson, Rainville, Rennie, & Mueller, 2016), and “citizen science” methods of getting more spatially distributed observations (e.g., Le Coz et al., 2016; Paul et al., 2018; Starkey et al., 2017). Significant epistemic uncertainties and some unmeasured states remain for all of these technologies. For some variables, the uncertainties might be reduced; for others, it might be necessary to seek new methods. There also remain important questions to be resolved about just how to test models as hypotheses when there are important epistemic uncertainties in the observational data, but certainly, a good starting point would be to reduce those uncertainties as far as technologically possible. This is likely to require some radically new approaches to provide a step-change improvement, given the limitations of existing observational techniques.

For such long-term aims, we might draw an analogy with defining a new satellite system for Earth Observation, such as the SWOT mission (e.g., Biancamaria et al., 2009; Biancamaria, Lettenmaier, & Pавelsksy, 2016). First, we need to define a functional requirement and then a technical specification and provide a justification for funding, including simulations of the difference the sensor would make, before any satellite-based sensor can be designed, built, and successfully launched. SWOT was listed as a potential mission in NASA's Decadal Plan of 2007; it will hopefully be launched in 2021. In the meantime, SWOT work has generated a large number of papers about how the data will contribute to improving estimates of the global water balance, flood discharges and inundation from larger rivers, surface storage in lakes, and the calibration of hydrological models.
As hydrological science moves into the future, it seems essential to improve observational methods in testing process representations and thereby gaining improved understanding. The British Hydrological Society Working Group suggested a number of long-term needs for improved observational methods (to download the full report, including suggestions on shorter term needs and model and theoretical developments, go to http://www.hydrology.org.uk/bhs-working-group-future.php):

- discharge measurements sufficiently accurate to calculate incremental discharges downstream;
- catchment precipitation inputs to much higher accuracies for better characterization of catchment water balance and forecasting purposes;
- total subsurface storage at scales useful for defining some "process response unit";
- better characterization of dynamic storages in different layers; and
- better characterization of controls on fluxes of water and solutes in different layers (including hot spots/hot moments/preferential flows/non-homogenous turbulence/... ) in relation to soil hydrological functioning and land management.

A combination of such field observations and model testing might be one way of combatting the general decline of field hydrology relative to modelling (e.g., Burt & McDonnell, 2015). In doing so, however, we need to be ambitious: to start to evaluate just where the biggest advances might be made for the purposes of both hydrological science and applied hydrology. Initially, this would have to make use of the type of prior simulations suggested earlier, testing how different levels and types of observation might make a difference to hypothesis testing and hydrological practice. These combinations should lead, as a community effort, to defining and commissioning new technologies and would, we believe, lead to significant gains for hydrological science. There is, of course, the question of who would pay for those new technologies to be developed and made available, which also depends on issues of who might invest and who benefits, but the important point is that we should make a start on deciding what should be prioritized, even if the process might be long term.

Jim Freer https://orcid.org/0000-0001-6388-7890
David M. Hannah https://orcid.org/0000-0003-1714-1240
Kate Heppell https://orcid.org/0000-0001-6028-1359
Joseph Holden https://orcid.org/0000-0002-1108-4831
Rob Lamb https://orcid.org/0000-0002-9593-621X
Thorsten Wagener https://orcid.org/0000-0003-3881-5849

REFERENCES

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., ... Plont, S. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, 12, 533–540. https://doi.org/10.1038/s41561-019-0374-y

Bashford, K. E., Beven, K. J., & Young, P. C. (2002). Observational data and scale dependent parameterisations: Explorations using a virtual hydrological reality. *Hydrological Processes*, 16(2), 293–312. https://doi.org/10.1002/hyp.339

Ben-Zvi, M., Berkowitz, B., & Kesler, S. (1988). Pre-posterior analysis as a tool for data evaluation: Application to aquifer contamination. *Water Resources Management*, 2(1), 11–20. https://doi.org/10.1007/BF00421927

Beven, K. (2019a). Towards a methodology for testing models as hypotheses in the inexact sciences. *Proceedings of the Royal Society A*, 475(2224), 20180862. https://doi.org/10.1098/rspa.2018.0862

Beven, K. J. (2006). The Holy Grail of scientific hydrology: as closure. *Hydrology and Earth Systems Science*, 10, 609–618. https://doi.org/10.5194/hess-10-609-2006

Beven, K. J. (2012). *Rainfall–runoff modelling: The primer (2nd ed.)*. Wiley Blackwell: Chichester, UK.

Beven, K. J. (2019b). How to make advances in hydrological modelling. *Hydrology Research*, in press. https://doi.org/10.2166/nh.2019.134

Beven, K. J., & Germann, P. F. (2013). Macropores and water flow in soils revisited. *Water Resources Research*, 49(6), 3071–3092 https://doi.org/10.1002/wrcr.20156

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., ... Watson, M. (2018). Epistemic uncertainties and natural hazard risk assessment—Part 2: What should constitute good practice? *Natural Hazards and Earth System Sciences*, 18(10), 2769–2783. https://doi.org/10.5194/nhess-18-2769-2018

Beven, K. J., & Lane, S. (2019). *Invalidation of models and fitness-for-purpose: A rejectionist approach*, Chapter 6. In C. Beisbart, & N. J. Saam (Eds.), *Computer simulation validation—Fundamental concepts, methodological frameworks, and philosophical perspectives* (pp. 145–171). Cham: Springer.

Biancamaria, S., Andreassis, K. M., Durand, M., Clark, E. A., Rodriguez, E., Mognard, N. M., ... Oudin, Y. (2009). Preliminary characterization of SWOT hydrology error budget and global capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1), 6–19.

Biancamaria, S., Durand, M., Andreassis, K. M., Bates, P. D., Boone, A., Mognard, N. M., ... Clark, E. A. (2011). Assimilation of virtual wide swath altimetry to improve Arctic river modeling, *Remote Sensing of Environment*, 115(2), 373–381. https://doi.org/10.1016/j.rse.2010.09.008

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. In *Remote Sensing and Water Resources* (pp. 117–147). Cham: Springer. https://doi.org/10.1007/978-3-319-32449-4_6

Birkel, C., Hellwell, R., Thornton, B., Gibbs, S., Cooper, P., Soulsby, C., ... Midwood, A. J. (2018). Characterization of surface water isotope spatial patterns of Scotland. *Journal of Geochemical Exploration*, 194, 71–80. https://doi.org/10.1016/j.jgeexplo.2018.07.011
