@AGUPUBLICATIONS

Water Resources Research

COMMENTARY

10.1002/2013WR015141

Key Points:

- Celerity is faster than velocityWatershed models are evaluated
- only against celerityVelocity must be addressed in future watershed models

Correspondence to: J. J. McDonnell,

jeffrey.mcdonnell@usask.ca

Citation:

McDonnell, J. J., and K. Beven (2014), Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities, and residence time distributions of the headwater hydrograph, *Water Resour. Res.*, *50*, 5342–5350, doi:10.1002/ 2013WR015141.

Received 3 DEC 2013 Accepted 15 APR 2014 Accepted article online 19 APR 2014 Published online 2 JUN 2014

Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph

Jeffrey J. McDonnell^{1,2,3} and Keith Beven^{4,5,6}

¹Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, ²School of Geosciences, University of Aberdeen, Aberdeen, Scotland, ³Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, Oregon, USA, ⁴Lancaster Environment Centre, Lancaster University, Lancaster, UK, ⁵Department of Earth Sciences, Geocentrum, Uppsala University, Uppsala, Sweden, ⁶Centre for the Analysis of Time Series, London School of Economics, London, UK

Accurate prediction of the headwater hydrograph implies adequate modeling of sources, flowpaths and residence time of water and solutes.

Hewlett and Troendle [1975]

1. Introduction

Headwater, or first-order, catchments are the building blocks of drainage basins. They contribute up to half of the mean water volume and nitrogen fluxes for fourth-order and higher-order rivers and navigable waters in the USA [*Alexander et al.*, 2007]. While headwaters have been studied intensively since the First International Hydrological Decade (IHD) in the 1960s, major advancements in understanding have been rare. Perhaps, one of the most fundamental in the post IHD-era is the recognition that the mean transit time for water through headwater catchments can exceed the time scale of hydrologic response by orders of magnitude [*Martinec*, 1975]. Or, as *Kirchner* [2003] notes, that stream discharge responds promptly to rainfall or snowmelt, but with flows comprised of water that can be years or decades old.

Of course, differences between the flow velocities in the system (that control the tracer response) and the celerities (or speed with which perturbations are transmitted, which control the hydrograph) are to be expected since they are controlled by different mechanisms. Nevertheless, for years we have been trying to reproduce the way in which a catchment responds to precipitation through rainfall-runoff models tested against only one of these measures (the hydrograph). This focus has been treated as a problem of rainfall-runoff model parameter identification against values of stream discharge through optimization. This line of work continues through the use of multicriteria Pareto optimization and use of Bayesian analysis for identification of well-constrained posterior parameter distributions—all related to discharge measurements. But this is only half the story in properly understanding how water moves through catchments.

More recently, it has been suggested that many different models might provide acceptable reproduction of the hydrograph (the equifinality thesis of *Beven* [2006a, 2012b]) and that model identification might then be best posed in a framework of testing models as multiple working hypotheses about catchment response. In testing models as hypotheses, we want to get the right results for the right reasons [*Kirchner*, 2006] but this is difficult when there are errors, uncertainties, and inconsistencies in the precipitation input and discharge data time series. More problematic, and not fully appreciated in the literature to date (but certainly noted by *Hewlett and Troendle* [1975]), is that stream hydrographs alone have limited value in assessing if the model is working for the right reasons because they do not test adequately the processing of water within the catchment in both space and time.

Here we argue that perhaps the fundamental issue for improving our understanding of hydrological processes and our capability to model them is the explicit and routine use of celerities and velocities in model



development and testing. This will involve the joint collection and use of hydrographs and tracers to make inferences about both velocities and celerities within a catchment and explicit acknowledgement that accurate prediction of the headwater hydrograph implies adequate modeling of water sources, flow paths, and residence times. Admittedly, little instruction is given in our field in terms of how and what to measure to characterize water sources, flow paths, and residence times, as commented upon by *Seibert and Beven* [2009] and more recently by *Seibert and McDonnell* [2013]. What seems clear now is that any new watershed investigation should include the measurement of both hydrograph and tracer information, i.e., the collection of precipitation and runoff amount as well as the isotope composition of both signals at frequencies sufficient to characterize the changing nature of responses as a catchment wets and dries. While this short debate article lacks a thoroughgoing treatment of this matter, we attempt here to shine a light on this fundamental issue by clarifying the differences between velocities and celerities at the headwater scale where hillslope transmission (and not the stream channel itself) dominates the outlet response. We then make the case for the routine use of tracers in models and outline a community action plan for moving forward.

2. So What Is so Different Between Velocity and Celerity?

There is much confusion in the literature on what constitutes and differentiates velocity and celerity. Both words have their origins in the concept of speed; velocity from the Latin *velocitas* for speed, celerity from the Latin *celer* for swiftness. In contemporary usage, they have the units of volume flux per unit area (LT^{-1}) . For water, the word velocity is used to define the mass flux of the water itself. Celerity is defined with respect to the speed with which a perturbation to the flow propagates through the flow domain. The perturbation might be a pressure wave moving through the system, such as a flood wave moving down a river. This is why celerity is referred to as the wave speed in physics of fluids, fluid mechanics, and hydraulics (also known as the wave phase speed or frequency/wave number).

The difference between velocities and celerities, rather oddly, does not appear in most catchment hydrological text books, though related concepts such as "translatory flow" are sometimes referred to without clear explanation. The concept of wave translation goes back at least to the 1930s when it was discussed by Horton and Sherman and others [*Beven*, 2004]. However, most references to translatory flow cite the lateral subsurface stormflow work of *Hewlett and Hibbert* [1967] and the one-dimensional experiments of *Horton and Hawkins* [1965] that demonstrated the displacement of stored water (though note that those experiments had time scales of weeks, not hours)

The difference between velocities and celerities is most easily visualized for the experimental case used by Henri Darcy of a cylinder full of sand and saturated with water. Changing the flow rate or head at the input boundary will immediately cause a change in flow at the output boundary. While the water flow velocity through the sand is slow, the celerity in this case is (theoretically) instant, hence the immediate response. At larger scales, this case is analogous to a confined aquifer with incompressible water and rock. Allowing for the compressibility will slow the celerity a little (see Table 1), but the velocities of flow will still be much less than the celerities.

A more interesting case is that of a sloping soil box that is only partially saturated, with a water table maintained by a constant head upper boundary. Changing that boundary will cause a perturbation to move downslope with a celerity that depends on the magnitude of the change and the profile of water content in the unsaturated zone. In this case, the soil will generally be at or near saturation in the capillary fringe that exists above the water table so that small perturbations will have high celerities in the fringe; larger perturbations will have to fill or drain more pore space so will have lower celerities. Note that in a homogeneous soil box, the velocities of flow would not be changing very much at all, at least if the changes did not greatly change the slope of the water surface or if the change in hydraulic conductivity with change in water table was not highly nonlinear. At larger scales, this case is analogous to a saturated zone in the soil on a hillslope.

If we can assume in a headwater catchment that the changes to the slope of the hillslope groundwater surface and the conductivity are small, we can use a kinematic analysis to define velocities and celerities in the saturated zone (*Beven* [2012a, pp. 174] and discussion of *Beven* [2006b]). In this case, there are two relevant definitions of velocity classically used by the hydrologist. The Darcian flux, v_D, is the volume flux of water per unit cross-sectional area of soil or

Table 1. Some Simple Relationships for Velocities and Celerities for Different Types of Flow [After Beven, 1989, 2012a; Rasmussen et al., 2000; Torres et al., 1998; Beven and Germann, 2013; Gonwa and Kavvas, 1986]^a

Type of Flow	Mean Velocity (LT^{-1})	Celerity (LT^{-1})	c/v
Free surface flow: St. Venant equation	v	$c = v \pm \sqrt{gy}$	(F ± 1)/F
Free surface flow: diffusion approximation (manning velocity equation in a wide channel)	$v = (1/n)y^{3/2}S_f^{1/2}$	c = 5v/3	5/3
Kinematic overland flow (power law)	$v = ay^b$	c = bay ^{b-1}	b/y
Incompressible confined saturated Darcian subsurface flow	$v_D = K_s d\phi/dx$	$c = \infty$	∞
Compressible confined saturated Darcian subsurface flow	$v_D = K_s d\phi/dx$	$c = v_D/S_e$	1/S _e
Unconfined saturated Darcian subsurface flow	$v_P = K_s dh/dx$	$c = v_D / \theta_e$	$1/\theta_{e}$
	$v_P = K_s dh/dx/\theta_s$		$\theta_{\rm s}/\theta_{\rm e}$
Kinematic subsurface saturated flow (homogeneous conductivity)	$v_D = K_s \sin\beta$	$c = v_D / \theta_e$	$1/\theta_{e}$
	$v_{\rm P} = K_{\rm s} \sin\beta/\theta_{\rm s}$		θ_{s}/θ_{e}
Kinematic subsurface saturated flow (power law conductivity profile)	$v_D = ah^b sin\beta$	$c = bah^{b-1} sin \beta/\theta_e$	$b/h\theta_e$
	$v_P = ah^b sin \beta/\theta_s$		$b\theta_s/h\theta_e$
Stokes law preferential flows (wetting)	$v = gf^2/3\eta$	c = 3v	3

^aMathematical terms are defined as follows: v = mean velocity; $v_D =$ Darcian velocity; $v_P =$ mean pore water velocity; y = mean depth of free surface flow; g = acceleration due to gravity; F = Froude number; n = Manning roughness coefficient; S^f is the friction slope; $\phi =$ piezometric head; h = mean water table height above impermeable layer; K = hydraulic conductivity; $\Phi =$ total hydraulic head; x = direction of flow; $S^e =$ effective storage coefficient; $\theta =$ soil water content; $\theta_s =$ saturated porosity of the soil; $\theta_e =$ effective storage deficit above water table per unit rise or fall of water table; $\beta =$ slope angle; a, b = power law parameters; f is film thickness; η is kinematic viscosity.

$$v_D = Q/A = K_s \sin \beta$$

Where Q is the mass flux of water, A is the area saturated orthogonal to the downslope direction of flow, K_s is the saturated hydraulic conductivity of the soil, and β is the slope angle with sin β approximating the local hydraulic gradient. The mean pore water velocity is the volume flux of water per unit cross-sectional area of pore space or

$$v_D = Q/(A \theta_s) = K_s \sin \beta/\theta_s$$

Since θ_s is necessarily less than 1, the mean pore water velocity is always faster than the Darcian velocity. It is the local mean velocity that is relevant to the transport of tracer through the soil. This will also be affected by the distribution of pore velocities around that mean.

The celerity, c, can be defined in a similar way as

$$c = K_s \sin \beta / \theta_e$$

where θ_e is an effective storage that is filled or drained per unit change in the water table height (also referred to as specific yield in shallow groundwater literature). The effective storage may change over time but since θ_e is necessarily less than θ_s , the celerity will always be faster than both the mean pore water velocity and the Darcian flux. If θ_e is small, for a small perturbation or a very wet soil, then the celerity might be much faster than the mean pore water velocity. This means the discharge response to a perturbation (such as a recharge event labeled with a tracer) can be much faster than the tracer response of that event which is controlled by the pore water velocities. It is, of course, quite possible that some local flow velocities in preferential flow paths might be as fast and might affect local celerities by filling local storage deficits, but the nature of the controlling mechanisms mean that velocities cannot be faster than celerities on balance.

It is important to note here the different mechanisms controlling velocities and celerities. Velocities are controlled by the characteristics of the filled water storage; celerities are controlled by the storage deficit that must be satisfied or remain as the saturated zone rises and falls. Similar relationships can be developed for other profiles of hydraulic conductivity [e.g., *Beven*, 1981]; for specific soil moisture characteristic curves [e.g., *Rasmussen et al.*, 2000]; and for preferential flows based on Stokes law [*Beven and Germann*, 2013] (see Table 1).

A similar simplified analysis can be made for the case of free surface water flows. In this case, for perturbations or waves of long wavelength and low amplitude, such as flood waves or in overland flow, the difference between celerities and velocities, v, is given by

$$c = v \pm \sqrt{gy}$$

where g is the acceleration due to gravity and y is the local mean flow depth. The symbol \pm indicates that for subcritical flow the perturbations will move in both upstream and downstream directions (think of the

perturbations caused by throwing a rock into a calm section of river). For supercritical flow ($v/\sqrt{gy} > 1$), there will be movement only in the downstream direction since the downstream flow velocity is greater than the speed by which the perturbation is traveling upstream.

The celerities in this case can be related directly to the characteristic curves of the governing depth averaged St. Venant equations. These are the trajectories in the space-time dimensions of constant values of depth or velocity. They are most easily visualized for the one-dimensional case. Subcritical flow has characteristic curves in both upstream and downstream directions; supercritical flow (and the kinematic wave approximation to the St. Venant equations) has characteristic curves only in the downstream direction. Similarly, for the Darcy-Richards equation for subsurface flow there can be characteristic curves in both upstream and downstream directions, but for the kinematic wave approximation discussed above there are characteristics only in the downslope direction. These characteristic curves are a means of tracking the movement of perturbations. They are interesting in that they can be used to reveal the potential for shock fronts to develop where characteristic curves moving in the same direction intersect. In catchment systems, however, such shocks are of more theoretical than practical interest, any potential for shocks is likely to be dissipated by the heterogeneities of soil and surface properties and the consequent variability in the velocities and celerities in space and time. Some of these relationships for velocities and celerities are summarized in Table 1.

Of course, the velocity-celerity issue becomes more interesting, and often more confusing in natural, preferential flow-dominated headwater catchments. A good example of this is the Maimai Debate (for review see McGlynn et al. [2002]) where Mosley [1979, 1982] conducted simple water-dye additions to short hillslope sections and saw rapid breakthrough of applied tracer through preferential flow paths, interpreted as the rapid throughflow of storm rainfall. These dye responses were 300 times faster than the measured saturated hydraulic conductivity of the soil. While not recognized as such by Mosley [1979], these dye breakthroughs represented the rapid tail pore velocity distribution and not the Darcy flux, as originally conceived. But future work at the site further confused the velocity-celerity issue. In their isotope-based response to the Mosley work, Pearce et al. [1986, pp. 1263] noted that "processes which deliver 'old' water are largely responsible for hydrograph generation" and go on to say "currently favored runoff mechanisms, which involve the rapid flow of 'new' water over the ground surface or through the soil matrix, soil macropores, or other rapid transit pathways, cannot explain the streamflow response." This "effusion confusion" relates to the celerity-velocity difference. The Pearce et al. [1986] and Sklash et al. [1986] isotope-based results, that seemingly contradict Mosley, were in retrospect a difference of measured local-scale pore water velocities and the celerity component of flow. McDonnell [1990] showed how macropore flow of old water could explain these seemingly contradictory findings where transient water tables developing at the soil-bedrock interface could activate lateral, disconnected soil pipes to augment the old water effusion (as shown by Weiler and McDonnell, 2007, in a model analysis of the same site). Generally, process-based work in the intervening period has continued to show how connectivity of saturated patches in the subsurface (often aided by preferential flow at or near some subsurface permeability interface or contrast) is a prerequisite for threshold-like old water effusion from the slope base and into the stream or riparian zone [Bishop et al., 2004; McDonnell, 2013].

3. Toward the Routine Use of Tracers in Models

Our call to action is not, of course, totally new. There has been past encouragement to try and use other types of data within the model identification, evaluation and hypothesis testing framework. These have included the use of internal state data that provides additional information on the spatial patterns of the hydrograph response [e.g., *Franks et al.*, 1998; *Lamb et al.*, 1998; *Seibert and McDonnell*, 2002; *Blazkova et al.*, 2002; *Freer et al.*, 2004]. But the use of tracer data, with explicit recognition of the velocity-celerity differentiation, has been limited. One of the earliest uses of tracers in headwater catchment hydrology models was *Hooper et al.* [1988] who used geochemical and hydrologic information together in a rejectionist framework to examine model parameterization and parameter identifiability. *Harris et al.* [1995] used stable isotopes in an open system mixing model for the Gårdsjön F1 covered roof tracer addition experiment.

One of the best early examples of isotope time series representation was by *Seibert et al.* [2003] who developed a two-box model of catchment dynamics constrained by 18-O and hydrograph response together. Since then, several papers have examined tracer data together with hydrograph information to explore residence time dynamics of water within the catchment [e.g., *Vaché and McDonnell*, 2006; *Weiler and McDonnell*, 2007;

Sayama and McDonnell, 2009; Fenicia et al., 2010] with the work of Birkel et al. [2010, 2011], Hrachowitz et al. [2013], and Davies et al. [2011, 2013] illustrating the latest developments in this area. Collectively, these studies show clearly that if a model can reproduce the spatial pattern of response and the variability in residence times of the water over time, then we can have much more belief that we are getting the right results for the right reasons.

Perhaps the reason that there have not been many studies using tracers and hydrographs together is because neither tracer or spatially distributed state data are necessary for getting what appears to be an acceptable simulation of the hydrograph (or at least sufficiently acceptable to satisfy the review process). Indeed, using such data will normally require more complex models with more parameters to be identified. From a scientific point of view, however, it is important that we move forward from this situation. It is important that we subject our models to more rigorous testing, while recognizing the difficulties posed by data limitations in testing models as hypotheses. In particular, we would suggest it is particularly important that investment is made in combining mass flux measurements of the water with tracer flux measurements with sufficient detail in time to allow the time variability in the residence times of water to be assessed. There are very few such data sets available even for small research catchments. This needs urgent attention.

There are, however, also difficulties in using these data for model evaluation that relates to three factors: (1) issues of the commensurability of model and predicted variables, (2) the additional parameters that might be necessary to obtain acceptable predictions of the concentration data, and (3) the possibility that inconsistent input and output concentration data might introduce disinformation into the modeling evaluation process in the same way as for the mass fluxes of water [*Beven et al.*, 2011; *Beven and Smith*, 2014]. Commensurability issues will arise because of the space and time scales of sampling relative to those of the concentrations that a model is predicting.

It is important to note that introduction of additional parameters is necessary to allow for the additional mixing volume needed to predict water residence times relative to predicting the hydrograph. This difference has been ignored in some models of both hydrographs and tracer time series, when the storages used to predict the hydrograph have also been used to predict the tracer. This assumes, incorrectly, that celerities are equal to velocities. However, even adding additional parameters does not guarantee good reproduction of detailed tracer response [see, e.g., *Page et al.*, 2007].

4. Toward the Routine Collection of Water Tracer Data

Part of the problem with uptake of the routine use of tracers in models is that it has been logistically difficult and expensive to get good data on environmental tracers for sufficiently long periods of time to be really useful in model evaluation (although there is evidence that even sparse isotope data can be useful for a posteriori model rejection as shown by *Vaché et al.* [2004]). Detailed data sets have been limited largely to single storm events (see recent review on storm hydrograph separation by *Klaus and McDonnell* [2013]), short sequences of storm events [*lorgulescu et al.*, 2005, 2007], or a short period of high-frequency data [*Berman et al.*, 2009]. There has also been a tendency to collect bulk samples of event inputs, although it is known that precipitation concentrations can vary in time and space (and be subject to redistribution where snow is important).

But this situation is changing. For the environmental isotopes of oxygen and hydrogen that make up the water molecule, the increasing reliability of (relatively) cheap, off-axis and cavity ring-down laser spectrometers is revolutionizing the possibilities of collecting detailed data at reasonable cost. The priorities in collecting such data sets should be firstly getting much more detailed information in time. This has been demonstrated in the study of *Berman et al.* [2009] who collected data at 30 min intervals, but only over a 3 month period. It is also apparent in the wide range of geochemical species collected at Plynlimon at 7 h intervals by *Neal et al.* [2012] for a period of 18 months. The detail revealed in these data sets will be related, at least in part, to the time variability of the celerities and velocities, and hence the different residence time distributions of water in the system [*McDonnell et al.*, 2010; *Rinaldo et al.*, 2011] and to characteristics of the process controls that have not been revealed in coarser resolution data [e.g., *Harris and Heathwaite*, 2005; *Jordan et al.*, 2005; *Cassidy and Jordan*, 2011].

5. Summary and Outlook

We have argued here that the most important issue for improving our understanding of hydrological processes and our capability to model them is an action plan aimed at velocities, celerities and residence times of headwater hydrographs. Such a focus is especially needed in nonstationary times. Such a focus is also needed for getting the right answers for the right reasons. We still have much to learn about the joint simulation of both discharge and tracer concentrations, and additionally, more routine and good quality data sets from different types of catchments and in different environments are critically needed.

Field data needs. Environmental isotope tracers are likely to prove to be the most useful tracer for the foreseeable future at headwater catchment scales due to logistical issues of applied tracers beyond the hillslope scale and the nonconservativeness of natural geochemical tracers. While laser spectrometers have made analysis of water isotopes fast and inexpensive, deployment into the field is still challenging given the power requirements, housing restrictions and instrument sensitivity to in situ use [*Pangle et al.*, 2013]. Work is required on multiple fronts. We need new work that enables field deployment of current laser spectrometer technologies. This includes the need for new devices to route rainfall and streamflow into current generation laser spectrometers and the development of nebulizers to convert liquid water to vapor to enable laser-based vapor spectrometers to be used in the field. But we should not await the widespread availability of cheap, field deployable laser specs before we begin the routine collection of precipitation and streamflow isotope samples. Instead, we should begin that collection now with manual and automatic liquid samplers to start to amass the important data necessary for model analysis. New watershed studies should commit to collect such water samples as normal practice wherever possible.

Process representation needs. Consideration of the difference between velocities and celerities in catchment responses leads to the conclusion that process representations will need to be scale dependent and hysteretic [e.g., *Beven*, 2006b, 2012a, Chapter 9; *Rinaldo et al.*, 2011]. In doing so, however, it will be necessary to deal not only with the differences between celerities and velocities but also the way in which they change with space and time and connectivity. New theoretical developments are needed to match the types of observations that will become available [e.g., *Davies et al.*, 2011, 2013]. Perhaps the most important aspect that combined velocity and tracer studies can contribute to understanding is the explicit description of the dynamics of watershed storage. Knowing the hydrograph and tracer response will allow much better hypothesis testing of different representations of how water is processed through the catchment storage.

Model evaluation. Models can be both tools of convenience when we require predictions of catchment systems for management of water resources, floods, or water quality, or ways of formulating hypotheses about system functioning. Ideally, of course, demonstrating that we have good hypotheses would feed back into the management tools used in practice. It is not clear that this is happening, and this is in part due to the fact that it is actually not very difficult to get a model to (more or less) fit a sequence of hydrographs. This is especially true if we make allowance for the uncertainty intrinsic to both model inputs and the discharges with which a model can be calibrated or evaluated. We need more stringent tests of models, both for science and for practice, while taking account of such uncertainties. There are various methods for model evaluation and this remains a topic of some rigorous debate [e.g., *Clark et al.*, 2011; *Beven et al.*, 2012]. In that such evaluations will necessarily involve epistemic as well as aleatory uncertainties [e.g., *Beven*, 2012b; *Beven and Young*, 2013], progress in model evaluation will depend on applications involving new types of testing against new types of data sets.

While we recognize the real difficulties of collecting high-frequency and high-quality data sets of this type (through past experiences of trying to do so), we suggest that this should be one of the highest priorities in future hydrological research projects. Because of the different mechanistic controls on celerities and velocities, the joint use of hydrograph and concentration data will reveal more about the system response, even if models might require more parameters to be defined in simulating the processes involved. The time is right. Let us not want to wait another 40 years to agree that accurate prediction of the headwater hydrograph implies adequate modeling of sources, flow paths and residence times.

6. Postscript: Comments on the Other Papers in the Debate

There is now a recognition in the hydrology community that the use of models is an exercise in constraining prediction uncertainty. While there is not always agreement on how that uncertainty should be assessed in evaluating model predictions [*Clark et al.*, 2011; *Beven et al.*, 2012] this recognition applies to both global-scale hydrology and to catchment-scale predictions as noted in the two debate submissions in this volume [*Lall*, 2014; *Gupta and Nearing*, 2014]. The problem is that regardless of scale, the data we have available is not usually good enough to distinguish between different hypotheses about how catchment systems

function [e.g., *Beven and Westerberg*, 2011; *Kauffeldt et al.*, 2013; *McMillan et al.*, 2012; *Beven and Smith*, 2014]. Consequently, there is really not a clear difference between the equifinality concept of *Beven* [2006a, 2006b, 2012a, 2012b] and the diversity of hypotheses approach suggested by *Gupta and Nearing* [2014]. If, given a diversity of hypotheses, we have the data that enables rejection of some hypotheses, then we should of course do so. However, if we cannot distinguish between the hypotheses, then we have a degree of model equifinality that cannot yet lead to a hypothesis being rejected. There seems therefore to be very little difference between the concepts in this instance.

Data is key. But what data is best? Our proposal is to focus on tracer data to test models in fundamentally different ways. In fact, it links to the broader and perhaps most important issue of all in hydrology: epistemic uncertainty (i.e., knowledge of how the system works) in our understanding of the processes and mechanisms of water cycling in catchments during and between events. As modelers, we are often overly confident in our perceived knowledge of the states, stocks and flow paths of water and pay little attention to residence times. As experimentalists, we recognize how complex catchment systems are and how impossibly difficult they are to characterize in detail. As we move to model structures that better capture the mechanisms of water cycling and the key thresholds, hysteresis, competitive feedbacks, storage effects and other nonlinearities that define them [*McDonnell*, 2013], we argue that acceptable model performance for both tracer and hydrograph observations can get us closer to getting the right results for the right reasons.

Some ambiguity or equifinality will surely remain, particularly when it is necessary to add parameters to make use of the tracer data. This does, however, illustrate more general points: (1) that we need better data sets and new observation methods to evaluate models more rigorously than in the past, and (2) that we need that data to be more informative in differentiating process representations at appropriate scales. The equifinality principle does not imply that we will never be able to distinguish between models as hypotheses – rather it suggests only that current data sets are not generally adequate to do so, particularly when the uncertainties are epistemic as well as aleatory in nature (i.e., the intrinsic randomness of the system), so that classical likelihood measures will lead to overconditioning [*Beven*, 2006a, 2006b, 2012a, 2012b]. But the point about epistemic uncertainties is that they should be reducible by improving our observational tools and knowledge. In the hydrology of the future we therefore need a much greater emphasis on new observational methods that will give us more tools for evaluating models as hypotheses.

Acknowledgments

These ideas have been brewing for many years and aided by useful discussions with Jan Seibert, Kellie Vache, Sarah Dunn, Doerthe Tetzlaff, Chris Soulsby, Markus Weiler, Fabrizio Fenicia, and Rick Hooper. More recent discussions with Laurent Pfister, Andrew Ireson, and Lee Barbour have been helpful in honing and sharpening ideas. Willemijn Appels, Chris Gabrielli, and Anna Coles are thanked for their reviews of an earlier version of this manuscript. The Global Institute for Water Security is thanked for providing funding for a short research visit where these ideas coalesced.

References

Alexander, R. B., E. W. Boyer, R. A. Smith, G. E. Schwarz, and R. B. Moore (2007), The role of headwater streams in downstream water quality, J. Am. Water Resour. Assoc., 43(1), 41–59.

Berman, E., M. Gupta, C. Gabrielli, T. Garland, and J. J. McDonnell (2009), High-frequency field deployable isotope analyzer for hydrological applications, *Water Resour. Res.*, 45, W10201, doi:10.1029/2009WR008265.

Beven, K. (1981), Kinematic subsurface stormflow, Water Resour. Res., 17(5), 1419–1424.

Beven, K., and P. Germann (1982), Macropores and water flow in soils, Water Resour. Res., 18(5), 1311–1325.

Beven, K., and P. Germann (2013), Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi:10.1002/wrcr.20156.
Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, Water Resour. Res., 49, 5092–5098. doi:10.1002/wrcr.20393.

Beven, K. J. (1989), Interflow, in Proceedings of NATO ARW on Unsaturated Flow in Hydrological Modelling, edited by H. J. Morel-Seytoux, pp. 191–219, Kluwer, Dordrecht, Netherlands.

Beven, K. J. (2004), Robert Horton and abrupt rises of groundwater, Hydrol. Processes, 18, 3687–3696.

Beven, K. J. (2006a), A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36.

Beven, K. J. (2006b), The Holy Grail of scientific hydrology: $Q_t = H(\underline{S}, \underline{R})A$ as closure, Hydrol. Earth Syst. Sci., 10, 609–618.

Beven, K. J. (2012a), Rainfall-Runoff Modelling: The Primer, 2nd ed., Wiley-Blackwell, Chichester, U. K.

Beven, K. J. (2012b), Causal models as multiple working hypotheses about environmental processes, C. R. Geosci., 344, 77–88, doi:10.1016/j.crte.2012.01.005.

Beven, K., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), Hydrol. Earth Syst. Sci., 15, 3123–3133, doi: 10.5194/hess-15-3123-2011.

Beven, K. J., and I. Westerberg (2011), On red herrings and real herrings: disinformation and information in hydrological inference, *Hydrological Processes*, 25, 1676–1680, doi:10.1002/hyp.7963.

Beven, K., P. Smith, I. Westerberg, J. Freer (2012), Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al., *Water Resour. Res.*, 48, W11801, doi:10.1029/2012WR012282.

Beven, K. J., and P. J. Smith (2014), Concepts of information content and likelihood in parameter calibration for hydrological simulation models, *J. Hydrol. Eng.*, doi:10.1061/(ASCE)HE.1943-5584.0000991, in press.

Birkel, C., D. Tetzlaff, S. Dunn, and C. Soulsby (2010), Towards a simple dynamic process conceptualization in rainfall-runoff models using multi-criteria calibration and tracers in temperate, upland catchments, *Hydrol. Processes*, *24*(3), 260–275.

Birkel, C., D. Tetzlaff, S. M. Dunn, and C. Soulsby (2011), Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models, *Water Resour. Res.*, 47, W02515, doi:10.1029/2010WR009547.

Bishop, K., J. Seibert, S. Köhler, and H. Laudon (2004), Resolving the double paradox of rapidly mobilized old water with highly variable responses in runoff chemistry, *Hydrol. Processes*, 18(1), 185–189.

Blazkova, S., K. Beven, P. Tacheci, and A. Kulasova (2002), Testing the distributed water table predictions of TOPMODEL (allowing for uncertainty in model calibration): The death of TOPMODEL?, *Water Resour. Res.*, *38*, W01257, doi:10.1029/2001WR000912.

Cassidy, R., and P. Jordan (2011), Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with nearcontinuous phosphorus time-series data, J. Hydrol., 405(1), 182–193.

Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res., 47*, W09301, doi:10.1029/2010WR009827.

Davies, J., K. J. Beven, L. Nyberg, and A. Rodhe (2011), A discrete particle representation of hillslope hydrology: Hypothesis testing in reproducing a tracer experiment at Gårdsjön, Sweden, Hydrol. Processes, 25, 3602–3612, doi:10.1002/hyp.8085.

Davies, J., K. J. Beven, A. Rodhe, L. Nyberg, and K. Bishop (2013), Integrated modelling of flow and residence times at the catchment scale with multiple interacting pathways, *Water Resour. Res.*, 49, 4738–4750, doi:10.1002/wrcr.20377.

Fenicia, F., S. Wrede, D. Kavetski, L. Pfister, H. H. G. Savenije, and J. J. McDonnell (2010), Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time estimation, *Hydrol. Processes*, 24(12), 1730–1742.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot (1998), On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res., 34(4), 787–797.

Freer, J., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining Dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, J. Hydrol., 291, 254–277.

Gonwa, W. S., and M. L. Kavvas (1986), A modified diffusion equation for flood propagation in trapezoidal channels, J. Hydrol., 83(1), 119–136.

Gupta, H. V., and G. S. Nearing (2014), Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, *Water Resour. Res., 50*, doi:10.1002/2013WR015096, in press.

Harris, G., and A. L. Heathwaite (2005), Inadmissible evidence: Knowledge and prediction in land and riverscapes, J. Hydrol., 304(1), 3–19.

Harris, D. M., J. J. McDonnell, and A. Rodhe (1995), Hydrograph separation using continuous open-system isotopic mixing, *Water Resour. Res.*, *31*(1), 157–171.

Hewlett, J. D., and A. R. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in International Symposium on Forest Hydrology, edited by W. E. Sopper and H. Lull, pp. 275–290, Pergammon, Oxford.

Hewlett, J. D., and C. A. Troendle (1975), Non-point and diffused water sources: A variable source area problem, in *Watershed Management*, pp. 21–46, American Society of Civil Engineers (ASCE).

Hooper, R. P., A. Stone, N. Christophersen, E. de Grosbois, and H. M. Seip (1988), Assessing the Birkenes model of stream acidification using a multisignal calibration methodology, Water Resour. Res., 24(8), 1308–1316.

Horton, J. H., and R. H. Hawkins (1965), Flow path of rain from the soil surface to the water table, Soil Sci., 100(6), 377–383.

Hrachowitz, M., H. Savenije, T. A. Bogaard, D. Tetzlaff, and C. Soulsby (2013), What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, *Hydrol. Earth Syst. Sci.*, *17*(2), 533–564, doi:10.5194/hess-17-533-2013.

lorgulescu, I., K. J. Beven, and A. Musy (2005), Data-based modelling of runoff and chemical tracer concentrations in the Haute-Menthue (Switzerland) research catchment, *Hydrol. Processes*, 19, 2557–2574.

lorgulescu, I., K. J. Beven, and A. Musy (2007), Flow, mixing, and displacement in using a data-based hydrochemical model to predict conservative tracer data, *Water Resour. Res.*, 43, W03401, doi:10.1029/2005WR004019.

Jordan, P., J. Arnscheidt, H. McGrogan, and S. McCormick (2005), High-resolution phosphorus transfers at the catchment scale: The hidden importance of non-storm transfers, Hydrol. Earth Syst. Sci. Discuss., 9(6), 685–691.

Kauffeldt, A., S. Halldin, A. Rodhe, C. Y. Xu, and I. K. Westerberg (2013), Disinformative data in large-scale hydrological modelling, *Hydrology* & Earth System Sciences, 17(7), 2845–2857, doi:10.5194/hess-17-2845-2013.

Kirchner, J. W. (2003), A double paradox in catchment hydrology and geochemistry, Hydrol. Processes, 17(4), 871–874.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W., X. Feng, and C. Neal (2000), Fractal stream chemistry and its implications for contaminant transport in catchments, Nature, 403(6769), 524–527.

Kirchner, J. W., X. Feng, and C. Neal (2001), Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations, J. Hydrol., 254(1), 82–101.

Klaus, J., and J. J. McDonnell (2013), Hydrograph separation using stable isotopes: Review and evaluation, J. Hydrol., 505, 47–64.

Lall, U. (2014), Debates—The future of hydrological sciences: A (common) path forward? One water. One world. Many climes. Many souls., Water Resour. Res., 50, doi:10.1002/2014WR015402, in press.

Lamb, R., K. Beven, and S. Myrabø (1998), Use of spatially distributed water table observations to constrain uncertainty in a rainfall–runoff model, Adv. Water Resour., 22(4), 305–317.

Martinec, J. (1975), Subsurface flow from snowmelt traced by tritium, Water Resour. Res., 11(3), 496-498.

McDonnell, J. J. (1990), A rationale for old water discharge through macropores in a steep, humid catchment, *Water Resour. Res.*, 26(11), 2821-2832.

McDonnell, J. J. (2013), Are all runoff processes the same? Hydrol. Processes, 27, 4103-4111, doi:10.1002/hyp.10076.

McDonnell, J. J., et al. (2010), How old is the water? Open questions in catchment transit time conceptualization, modelling and analysis, Hydrol. Processes, 24(12), 1745–1754.

McGlynn, B., J. J. McDonnell, and D. Brammer (2002), A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchment, New Zealand, J. Hydrol., 257, 1–26.

McMillan, H., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, *Hydrological Processes 26*(26), 4078–4111.

Mosley, M. P. (1979), Streamflow generation in a forested watershed, New Zealand, *Water Resour. Res.*, 15(4), 795–806.

Mosley, M. P. (1982), Subsurface flow velocities through selected forest soils, South Island, New Zealand, J. Hydrol., 55(1), 65–92.

Neal, C., et al. (2012), High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge, *Sci. Total Environ.*, 434, 3–12.

Page, T., K. J. Beven, and J. Freer (2007), Modelling the chloride signal at the Plynlimon Catchments, Wales using a modified dynamic TOP-MODEL, Hydrol. Processes, 21, 292–307. Pangle, L., J. Klaus, E. Berman, M. Gupta, and J. J. McDonnell (2013), A new method for in situ high frequency measurement of ²H and ¹⁸O in multiple water flows, *Water Resour. Res.*, 49, 7797–7803, doi:10.1002/2013WR013743.

Pearce, A. J., M. K. Stewart, and M. G. Sklash (1986), Storm runoff generation in humid headwater catchments: 1. Where does the water come from?, Water Resour. Res., 22(8), 1263–1272.

Rasmussen, T., R. Baldwin, J. Dowd, and A. Williams (2000), Tracer vs. pressure wave velocities through unsaturated saprolite, Soil Sci. Soc. Am. J. 64, 75–85.

Rinaldo, A., K. J. Beven, E. Bertuzzo, L. Nicotina, J. Davies, A. Fiori, D. Russo, and G. Botter (2011), Catchment travel time distributions and water flow in soils, *Water Resour. Res.*, 47, W07537, doi:10.1029/2011WR010478.

Sayama, T., and J. J. McDonnell (2009), A new time-space accounting scheme for to understand predicting streamwater residence time and hydrograph source components in catchments, *Water Resour. Res.*, 45, W07401, doi:10.1029/2008WR007549.

Seibert, J., and K. Beven (2009), Gauging the ungauged basin: How many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, 13, 883–892.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multi-criteria model calibration, Water Resour. Res., 38(11), 1241, doi:10.1029/2001WR000978.

Seibert, J., and J. J. McDonnell (2013), Gauging the ungauged basin: The relative value of hard and soft data, J. Hydrol. Eng., doi:10.1061/ (ASCE)HE.1943-5584.0000861.

Seibert, J., A. Rodhe, and K. Bishop (2003), Simulating interactions between saturated and unsaturated storage in a conceptual runoff model, *Hydrol. Processes*, 17, 379–390, doi:10.1002/hyp.1130.

Sklash, M. G., M. K. Stewart, and A. J. Pearce (1986), Storm runoff generation in humid headwater catchments: 2. A case study of hillslope and low-order stream response, *Water Resour. Res.*, 22(8), 1273–1282.

Torres, R., W. E. Dietrich, D. R. Montgomery, S. P. Anderson, and K. Loague (1998), Unsaturated zone processes and the hydrologic response of a steep, unchanneled catchment, *Water Resour. Res.*, 34(8), 1865–1879.

Vaché, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, Water Resour. Res., 42, W02409, doi:10.1029/2005WR004247.

Vaché, K. B., J. J. McDonnell, and J. P. Bolte (2004), On the use of multiple criteria for a posteriori parameter estimation, *Geophys. Res. Lett.*, 31, doi:10.1029/2004GRL021577.

Weiler, M., and J. J. McDonnell (2007), Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes, *Water Resour. Res.*, 43, W03403, doi:10.1029/2006WR004867.