OVERVIEW

Perceptual perplexity and parameter parsimony

Keith J. Beven [ID] | Nick A. Chappell [ID]

Lancaster Environment Center, Lancaster University, Lancaster, UK

Correspondence
Keith J. Beven, Lancaster Environment Center, Lancaster University, Lancaster LA1 4YQ, UK.
Email: k.beven@lancaster.ac.uk

Edited by: Stuart N. Lane, Editor-in-Chief

Abstract

This article reconsiders the concept of a perceptual model of hydrological processes as the first stage to be considered in developing a procedural model for a particular catchment area. While various perceptual models for experimental catchments have been developed, the concept is not widely used in defining or evaluating catchment models. This is, at least in part, because of the evident complexity possible in a perceptual model and the approximate nature of procedural model structures and parameterizations, particularly where there is a requirement for parameter parsimony. A perceptual model for catchments in Cumbria, North-West England, is developed as an exemplar and illustrated in terms of time varying distribution functions. Two critical questions are addressed: how can perceptual model hypotheses be tested at scales of interest, and how can constraints then be imposed on the basis of qualitative perceptual knowledge in conditioning predictive models? It is suggested that there is value in perceptual information, particularly in thinking about predicting the impacts of future change and that we still have much to learn about moving from observational and perceptual complexity to parsimonious predictability.

This article is categorized under:
    Science of Water > Hydrological Processes
    Science of Water > Methods

KEYWORDS

Cumbria, hydrological observables, model consistency, north-West England, process understanding

1 | THE DEVELOPMENT OF PERCEPTUAL MODELS FOR CATCHMENT RESPONSES

In the first edition of Rainfall-Runoff Modeling: The Primer (Beven, 2001a) a number of successive stages in the modeling process were distinguished. The first stage was considered to be potentially the most important. That was the specification of a *qualitative perceptual model* of how the catchment of interest might function. This concept was first outlined in the context of hydrological models by Beven (1987) who suggested that this would be something that would be personal to each hydrologist. This was expanded in Beven (2001a) as "*the perceptual model is the summary of our perceptions of how the catchment responds to rainfalls under different conditions or, rather, <u>your</u> perceptions of that response.*

A perceptual model is necessarily personal. It will depend on the training that a hydrologist has had, the books and articles he or she has read, the data sets that he or she has analyzed and, particularly, the field sites and environments of which he or she has had experience" (p. 3/4). It is therefore, a form of qualitative mental model that can incorporate far more complexity than what could, later in the modeling process, be represented in mathematical and computational form in order to make quantitative predictions.

The process of developing a perceptual model is a matter of experience. It should therefore, be expected to change over a time and to vary from person to person. There is a good example of this in the changing nature of the perceptual models for Maimai in New Zealand that has been recorded in the articles of Mosley (1982), Sklash et al. (1986), and Jeff McDonnell (Brammer & McDonnell, 1996; McDonnell, 1990; McGlynn et al., 2002). One example of how the developing perceptual model did feed into the structure of a simulation model is for the Ringelbach catchment in the Vosges, France (Ambroise et al., 1996; Freer et al., 1996). Perceptual models have also been proposed as the basis for model evaluation by Beven (1989a), Seibert and McDonnell (2002), and Vaché and McDonnell (2006) and for the development of hydrological classification schemes (e.g., Black, 1997; Sawicz et al., 2011; Wagener et al., 2007). This type of classification also underlies the definition of hydrological response units (HRUs) or hydrological similarity units (HSUs) in many semidistributed hydrological modeling frameworks and land surface parameterizations (albeit that this often only have a single conceptual structure with different parameter sets in the different HRUs).

In one sense the perceptual model can be considered as a set of hypotheses about hydrological functioning that should be tested as relevant for specific applications. Experiments can be designed to test specific process hypotheses, indeed the changing nature of experimentation at the Maimai catchment, particularly the use of isotope data, was fundamental to the change of perceptual understanding there. Such experimentation can also be exploratory, as in the type of abductive reasoning to create hypotheses about functioning discussed in Baker (2017). The use of dye tracer experiments at soil core, plot, and hillslope scales, albeit often destructive, have revealed the ubiquity of preferential flows in soils. That has resulted in preferential flows as an essential element of many perceptual models, but in this case without satisfactory incorporation into models that can be used to demonstrate importance at catchment scales (see Beven, 2010, 2018; Chappell, 2010; Jones, 2010). In groundwater systems, flows in fracture systems are also often part of the perceptual model for an aquifer model (particularly in karst systems) but direct observations of flows in rock fractures has often been limited to small scales, so aquifer modeling is often limited to integral conceptual approaches (e.g., Hartmann et al., 2014; Rahman & Rosolem, 2017) lacking evaluation against observed fracture flows.

A quick search of the wider scientific literature shows that perceptual models are used in a variety of different fields, particularly, of course, in psychology and cognition (e.g., Dosher & Lu, 2017;). In hydrology, the idea has been used quite widely, particularly in teaching, but is rarely formally taken into account when defining a predictive model (and we have seen plenty of past perceptual model failures when an analysis starts with an inappropriate procedural model, see discussion in Beven, 2020b, 2021). Even where modeling systems allow a wide variety of different components (e.g., MMS, CMF, SUMMA, SuperFLEX, and MARRMoT) there has been a tendency to be guided by goodness of fit, rather than any evaluation against a perceptual model (see the recent optimization of model structures in Spieler et al., 2020). SUMMA (Clark et al., 2015a; Clark et al., 2015b) has the stated aim of providing a unified framework for different process representations, noting that "*improvements in model fidelity require a sagacious choice of both process parameterizations and model parameters*" (Clark et al., 2015b, p. 2538) but does not provide any explanation as to how to be sagacious.

2 | PERCEPTUAL MODELS AND HYDROLOGICAL COMPLEXITY

The general lack of consideration of any explicit perceptual model is somewhat perplexing but there would appear to be a number of reasons why this might be the case. The first is the mismatch between bottom-up perceptual complexity and the simplicity of most hydrological models (and we include in that description the so-called physically based models which may be computational demanding but are still based on gross and incorrect simplifications of the physics, for example, Beven, 1989b, 2018; Beven & Germann, 2013). As a community we have avoided thinking deeply about how that particular conundrum might be resolved. In fact, when we know that we would ideally want models that are robust in calibration (and which should therefore, be parsimonious in their parameterization), it is far easier not to think about the problem at all, but rather take a more top-down approach to defining hydrological functioning as represented by model structures from observed behavior at the catchment scale (e.g., Chappell et al., 2006; Gnann

et al., 2021; Ockenden & Chappell, 2011; Sivapalan et al., 2003; Wagener et al., 2021; Young, 2003, 2013; Young & Beven, 1994). An example of such an approach is the study of Wrede et al. (2015) who defined different models within Super-FLEX for three subcatchments of the Attert catchment in Luxembourg where the geological characteristics suggested quite different perceptual models. This top-down approach is, however, somewhat limited by the long-standing problem of a lack of agreed ways of linking dynamic response characteristics (DRCs) identified from input–output behavior to specific hydrogeological classifications of catchments (Addor et al., 2018; Gnann et al., 2021; Kratzert et al., 2019).

If all we are interested in is reproducing past streamflow observations, then experience suggests there will be many model structures and parameter sets that will provide more or less equivalent fits to the available streamflow observations. Indeed, purely data-based and machine learning models might well provide better fits to data than any explicit parameterization of the component hydrological processes (Beven, 2020a; Kratzert et al., 2019). This has been referred to as the equifinality problem (see Beven, 1993, 2001b, 2006a; Beven & Freer, 2001). Furthermore, the number of acceptable models will increase if we start to take proper account of uncertainties in the hydrological observations, rather than optimizing models as if these data were certain (e.g., Beven, 2019; Beven & Smith, 2015).

While there are definitely applications where modeling similar conditions to those in past data might be useful, a more important question is what can we say about changed conditions in the future; not only in the inputs (rainfall, net radiation, etc.), but also in the catchment characteristics of vegetation, soils, and channel form? Will a calibrated model properly simulate the process changes expected under different climate forcing (even without a change in basin characteristics)? How should the (already uncertain) parameters be changed to reflect future land management effects or other changes in the catchment that may impact on the functioning of surface and subsurface flow processes (see, e.g., Buytaert & Beven, 2009)? Will a predictive model that is consistent with the perceptual model do better in an assessment of change than just using a model that gives the best fit to past data? There may also be a question of how these simulated changes, especially if large scale, might feedback into regional climate models. The land surface components of these climate models are still generally different from those used in catchment hydrological models, but these questions are beyond the scope of what we want to consider here and are the subject of continuing research elsewhere (e.g., Fowler et al., 2016, 2018).

3 | THE VALUE OF A PERCEPTUAL MODEL

Underlying these questions is the issue of whether we might be able to do better in simulation (past, present, and future) if we have a better perceptual model of the hydrology of a catchment area. There are certainly examples of where hydrologists have not distinguished themselves in this respect. In a couple of recent articles, the historical context of two perceptual models of catchment-scale rainfall to streamflow response is discussed, where the concepts have been decidedly befuddled over time. The first is the concept of time of concentration, classically defined as the time of a drop of water to flow from the furthest point of a catchment to the outlet (including in WMO, 1974) but which, in its practical application, should really be defined in terms of wave celerities (Beven, 2020b). The second is the infiltration theory of the 1930s and 1940s that has been applied widely in computer models, but which was realized even during this "*era of infiltration*" to have decidedly shaky foundations (Beven, 2021).

The infiltration theory, as a perceptual model, is usually assigned to Horton and his article of 1933. At the time, it was adopted with enthusiasm by others as a useful and more accurate tool for estimating "*surface runoff*" for engineering design purposes (e.g., Cook, 1946). Hydrologists of that time did not, of course, have the advantage of environmental tracer information that suggests event hydrographs often comprise of a large proportion of pre-event water, something that later caused a fundamental change in many perceptual models. They did, however, recognize that "*direct runoff*" to the hydrograph could be made up of an unknown mix of surface and subsurface flows. Horton (1942) got round this problem by suggesting that fast subsurface stormflow would likely be turbulent rather than a laminar subsurface flow and could be called "*concealed surface runoff*" and therefore analyzed in the same way. The perceptual reasoning of the era of infiltration therefore became somewhat circular, justifying the application of a procedural model useful for quantitative prediction but with (generally) an incorrect perceptual basis. It is difficult to say that we are not subject to similar circularities of reasoning in applications of models now (especially when fitting to past streamflow observations is allowable). Hence the importance of trying to be rigorous about defining and testing an appropriate perceptual model and using it to inform quantitative prediction. This can be extended to include not only knowledge about the processes, but also about the uncertainties in the modeling processes (e.g., Westerberg et al., 2017).

So in what way might a rigorous and realistic perceptual model be useful? There are a number of ways that can be suggested.

- To put more focus on the understanding of <u>observable</u> hydrological processes as a way of improving predictability.
- To provide some evaluation of the meaning and value of what is observable.
- To try harder to get the "*right results for the right reasons*."
- To evaluate predictive models for consistency with available hydrological understanding.

We will return to this issue of consistency after considering a particular perceptual model.

4 | A PARTICULAR PERCEPTUAL MODEL OF CATCHMENT RESPONSES IN CUMBRIA

So consider a perceptual model for the types of catchment in Cumbria, North-West England that are the subject of a current project on Natural Flood Management (i.e., nature-based solutions for reducing flood hydrographs). These gauged catchments range in scale from less than 1 km^2 to more than 2000 km^2. These catchments are mostly covered by grassland, with both rough upland and improved pasture, with small areas of cultivation in the larger valleys. They have steep slopes in the headwaters, but the depth of subsurface flow varies widely from being confined to the topsoil to deep flows within permeable rocks. Glacial till (diamicton) is present in most catchments, which has led to the development of gleyed soils and perched water-tables where they overlay permeable rocks. The climate is humid temperate, with average rainfalls ranging from 1,000 to 4,000 mm/year. There is currently an extensive programmer of tree planting, and other natural flood management strategies such as leaky dams and off-line storages being deployed, some with new monitoring at the micro-catchment (<1 km^2) scale and "feature-scale."

So here is a situation where it would be useful to be able to describe the response under current conditions and also what might be the case under changed (post-intervention) conditions; to permit effective decision making as to how best to invest in future, more widespread mitigation strategies. Consider what a perceptual model of the processes of catchment response might look like for these types of catchments. The following might be considered some of the key elements of such a perceptual model (PM) for these catchments.

PM1: The structure of the vegetation, and its seasonal changes, might be important in controlling the volume of water that reaches the ground surface. This will be through rainfall interception by the canopy and subsequent wet-canopy evaporation; and the effects on snow accumulation and albedo during melt periods. In windy areas, such as is often the case for large storms in North-West England, evaporation from rough, wet, canopies can be very sensitive to even small local humidity deficits resulting in the potential for some surprisingly large rates of loss (Page et al., 2020). There will also be a distribution of rainfall intensities associated with the type of rainfall event, and the redistribution of the net rainfall by the local vegetation canopy. Particularly for deciduous woodland these factors will change with season. While the throughfall component will, in general, have a lower average intensity than the rainfall above the canopy, stemflows might lead to local concentrations of input that might be important in exceeding local infiltration capacities or inducing bypassing flow in the soil (e.g., Johnson & Lehmann, 2006; Tischer et al., 2020).

PM2: That variability in input intensities will interact with surface conditions to affect infiltration. Robert Horton suggested that infiltration capacities of the topsoil were controlled by surface processes (compaction by raindrops and traffic, mobilization of fine particles that would block larger voids, etc., see Beven, 2004a) and that this also meant that we should expect infiltration capacities to vary over time, both seasonally and between events (see also Chappell & Lancaster, 2007, in North West England). Horton argued strongly that infiltration is not generally profile controlled, even though his equation has been shown to be equivalent to a solution to the Darcy–Richards equation after making some specific assumptions about the soil moisture characteristics (e.g., Eagleson, 1970). Temperature itself will have an effect on infiltration capacities through changes in water viscosity. Changes due to agricultural operations might be more significant in some of the arable areas of the larger catchments considered here, perhaps less so for the established grass covered areas. Zaslavsky and Sinai (1981) also discussed the possibility of poorly measured layering close to the soil surface and below leading to downslope flow in unsaturated soils, referred to since as a "*thatched roof effect*". There may be connectivity thresholds for overland flow pathways, including effects of runoff–runon phenomena (Bonell & Williams, 1986) and feedbacks between depths of overland flow and infiltration rates (Dunne et al., 1991). In some circumstances, infiltration rates might be limited by air

pressure effects (e.g., Dixon & Linden, 1972). Horton already recognized the importance of cracks and macropores in allowing the infiltration of water and the escape of air to the surface (Beven, 2004a).

PM3: Where overland flow, by either infiltration excess or saturation mechanisms, is produced, the concentration of water in micro-rills will be more efficient at shedding water downslope. Some of the slopes in the micro-catchments of interest here have natural pipes in areas of peat soils. These flowlines are evident from depression lines with collapse holes and in the vegetation patterns at the surface, with species indicators of wetter conditions (e.g., the soft rush, *Juncus effuses*). Models often make the convenient assumption that overland flow can be treated as a sheet flow. This also dates back to Robert Horton who, following a suggestion of Leroy Sherman, made use of a representation of a unit width hillslope in order to apply the Manning equation to overland flow, with the hydraulic radius approximated as the slope width (Horton et al., 1934). This allowed Horton to make his theoretical advances in relating overland flow to erosion in his seminal 1945 article, but the sheet flow assumption is not very realistic on natural surfaces (Emmett, 1978). In some cases the channelling might be the result of co-evolutionary processes (e.g., the piping that can develop in dispersive soils) or external forcing, such as the piping in some upland peat soils that is thought to be initiated by desiccation cracking in rare dry summers (Gilman & Newson, 1980; Jones, 2010).

PM4: Once water has infiltrated into the soil then the pattern of flow through the unsaturated zone is likely to be highly variable, as a result of variability in initial soil moisture storage, heterogeneity in soil characteristics, the potential for fingering or Stokesian film flow rather than capillary potential driven wetting fronts, and the potential for flow through macropores. The macropore pathways will be discontinuous, which might lead to local zones of saturation and displacement of stored water into other faster pathways. There is also evidence that fingering and film flows can also induce exchanges with and displacement of stored water, and that given a sufficient depth of soil and unconsolidated materials, drying fronts might overtake local wetting following the end of rainfall, before the saturated zone is reached. Not all flow through the unsaturated zone will be vertical, there may be downslope flows according to the structure and macroporosity of the soil. We would expect capillary forces to be important in retaining water in the soil matrix that might later be used in transpiration, while capillary rise might be important in sustaining water supply to roots in some cases. We might also expect that roots might grow toward more readily available water more quickly than water can move to the roots under dry conditions. Exchanges with relatively immobile water might be important in controlling the travel time distribution through the unsaturated zone.

PM5: In the saturated zone, there may be local areas of perched and local saturation that may be disconnected from the stream. These may be above the till or in zones of higher permeability bedrock and might be important sources of water to trees under dry conditions. This threshold-type connectivity may in part depend on the form of the hillslope; convergent slopes will be more likely to stay connected for longer, and perhaps produce saturated areas at the surface than divergent slopes (another type of connectivity threshold). Subsurface preferential flow pathways might not follow the surface topography. This is particularly true in the limestones of North West England where karst features may also lead to transfers of water across catchment divides defined by surface topography. Again, not all water in the saturated zone is expected to be readily mobile; relatively immobile water may affect the travel time distribution of water elsewhere in the system.

PM6: Return flow to saturated areas and the stream riparian zone might be important, particularly in areas with significant valley bottom infill, or where there are returning contributions from the permeable solid geology. Areas of exfiltration from the subsurface to the stream might be subject to significant variability in space and time resulting from heterogeneities in permeability characteristics in the immediate zone around the channel (including fracture zones in the underlying geology and, at larger scales, the structure of floodplain sediments), as shown, for example, by Käser et al. (2014) in North West England. It has long been recognized that the near-stream areas can be an important source of overland flow from exfiltrating water, especially where the hillslopes absorb a high proportion of event water (e.g., Dunne & Black, 1970; Hewlett & Hibbert, 1967). It has also long been recognized that storage within and on flood plains can be an important control on hydrograph shape (e.g., see Beven, 2004b). Less widely considered is the effectiveness of the exchanges between floodplains and rivers. It is known that the patterns of subsurface inputs to streams can be highly variable, particularly where a hyporheic zone is present (Alley et al., 2002).

PM7: Catchment systems are not always underlain by impermeable bedrock (and even "impermeable" bedrocks might allow a degree of deeper seepage in secondary porosity and fracture systems). Deep flows in the system might also have an effect on water balance where the catchment is very small or where the groundwater divide is different from the surface divide. This was demonstrated at one of the early very detailed studies of soil moisture profiles and groundwater levels at the Whitehall catchment in Georgia by Tischendorf (1969, see Figure 1.6 in Chorley, 1978). It can also lead to gauge underflows in valley bottom deposits and bedrock. This was perceived as affecting discharges at one of the Panola gauging stations, for example. Simulations suggest that even small amounts of deeper seepage might constrain the build-up of a saturated zone in the soil profile and limit downslope flow connectivity (e.g., Ahuja & Ross, 1983).

PM8: In the channel network, flow processes will be affected by pool, fall, riffle and dead zone sequences; downstream variability in channel and floodplain geometry; local modifications in channel geometry during and between events; seasonal variations in head loss due to macrophyte growth; and sediment concentrations. Ephemeral and discontinuous stream channels will reflect the local balance of inputs and outputs in reaches of the channel network.

PM9: Increases in scale introduce different types of behavior and range of processes. We expect that there will be a wider range of geology, soil and land cover types, with channel routing and floodplain storage becoming more important relative to hillslope hydrographs as we increase catchment scale.

Each of these brief descriptions could, of course, be expanded with more examples from specific catchments in North West England or elsewhere. We could, indeed, build up an evidence base for the type of perceptual model we would expect in different catchment settings within a study network. We could also clearly add to the list if we were also interested in other hydrologically driven features of the catchment system related to aquatic ecology, carbon cycling, or nutrient and sediment transport. The model outlined above is perhaps rather general in that it might apply to water fluxes in a rather wide range of sloping, humid temperate experimental catchments around the globe (e.g., Maimai, Panola, Andrews Forest, Plynlimon, etc.) but that is partly because of our joint experience in visiting and working in some of those catchments and reading what others have had to say about them. What has not generally been evident is how such a perceptual model can be linked to a mathematical description of the processes and then how that description might be tested as a series of research hypotheses.

A schematic outline of such a perceptual model is represented in Figure 1 with components labeled as in the text above. An interpretation of that schematic model in the form of a cascade of multiple distribution functions in shown in Figure 2 (as modified from that in Beven, 1989a). Note that the figure does not reflect the dependencies between distributions (e.g., between celerity and distance from the stream) or the way that they may be non-stationary in time. This, already simpler schematic model, immediately illustrates the problem of increasing numbers of parameters as more complexity is introduced, even if at least some of those distributions could be considered stationary. Moving from perceptual models to even simpler conceptual storage models is also illustrated in Wrede et al. (2015). These models
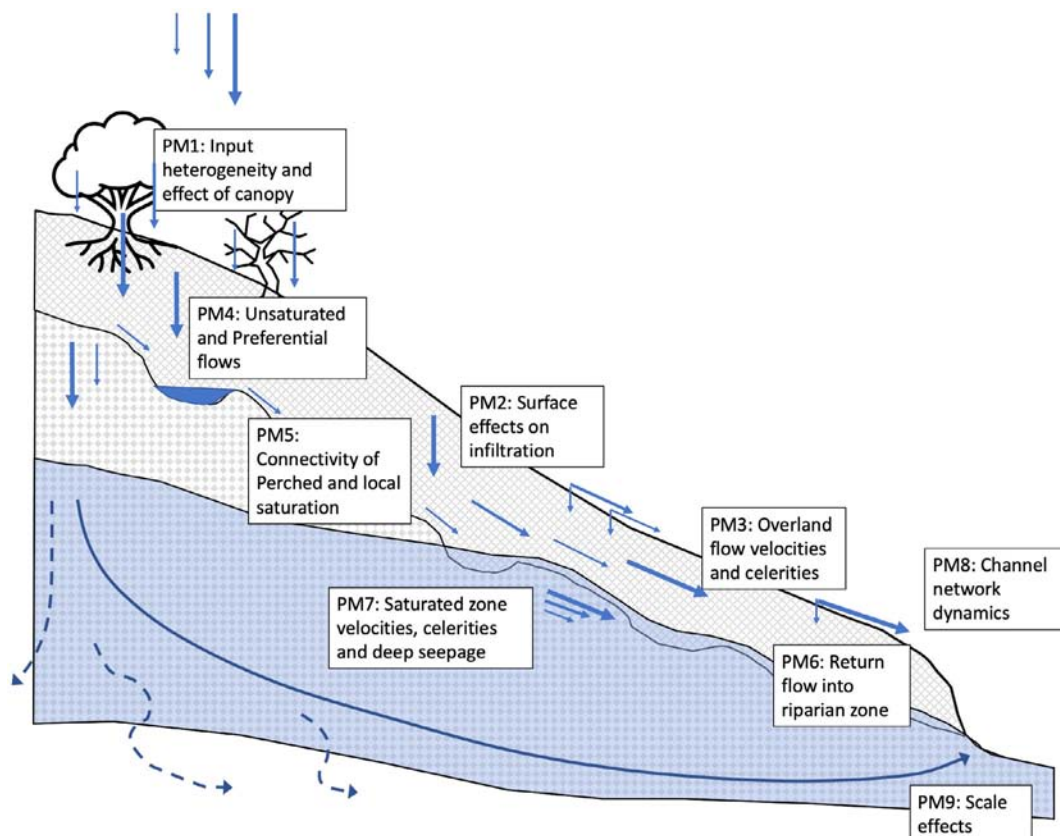

FIGURE 1 A schematic representation of the perceptual model with components as outlined in the text

Type of Distribution

PM1a: Pattern of intensity (i) above
canopy or surface
PM1b: Processing of intensities by the
canopy as throughfall (tf) and stemflow
(sf)
PM2: Pattern of surface infiltration
capacities (f_c)
PM3: Distribution of overland flow
celerities, (c_{uz})
PM4: Vertical soil water celerities
including preferential flows (c_{uz})
PM5: Perched saturation and storages
to be filled to allow downslope
connectivity (S_c)
PM6: Saturated zone wave celerities
including preferential flows and return
flows to saturated areas and riparian
zone
PM7: Deep seepage celerities (c_{ds})
including within fractures
PM8: Channel routing dynamics and
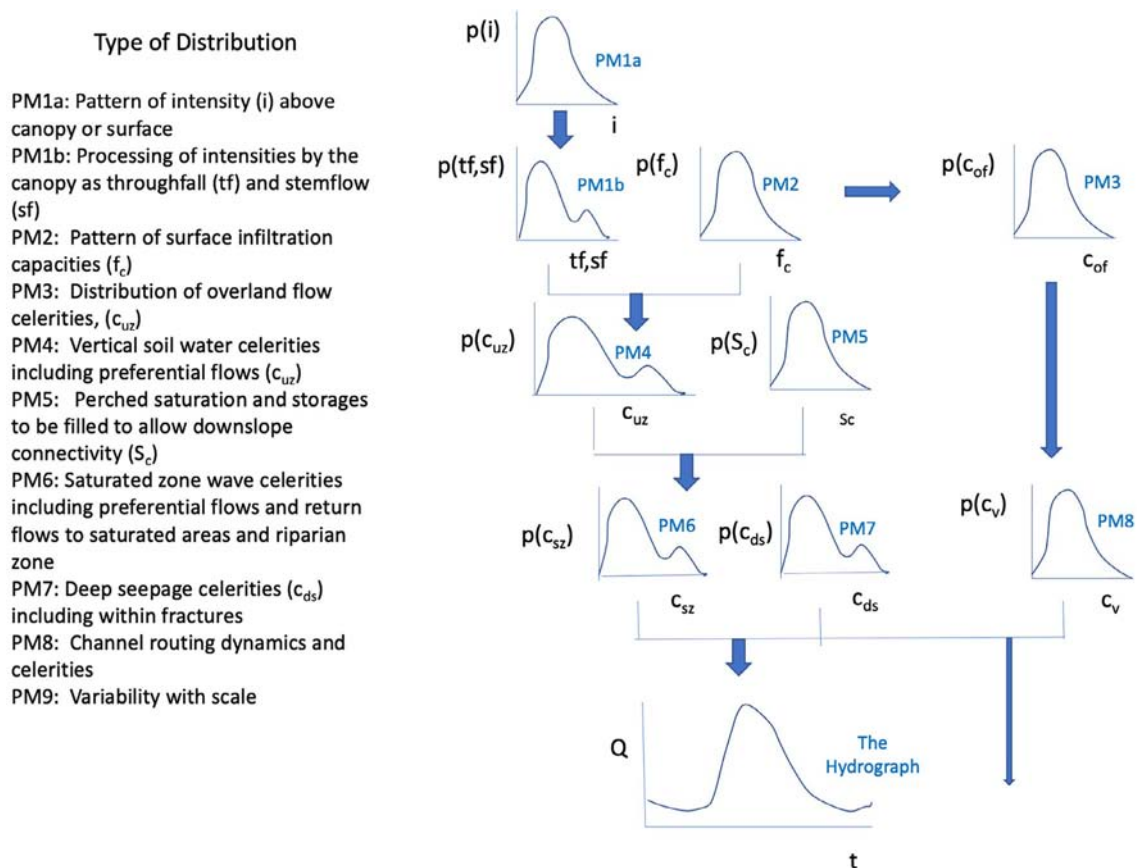celerities
PM9: Variability with scale

FIGURE 2 A representation of a perceptual model in the form of a cascade of distribution functions (modified from Beven, 1989a)

however, are tested only in the sense that the calibrated model structures are generally shown to perform better on the catchment with a corresponding perceptual model than for catchments with different perceptual models (particularly in the need for a deeper subsurface storage). This is not really the type of consistency with observable behaviors we are proposing here since the behavior is in essence being derived from the hydrographs that the model is being calibrated to. It would be good to have additional independent observations for model evaluation and hypothesis testing, even if that evidence might be necessarily uncertain (Beven, 2015, 2019).

Describing a perceptual model is not so difficult, but we would suggest that many hydrological modelers have become too remote from field experience to have credible local examples of processes operating at a particular site if only because of the limited resources often available to practical modeling applications. Certainly, it seems that few actually think very hard about this perplexing problem of representing the essential elements of that perceptual model in a working procedural model. This is surely partly because there have been only few exhortations to do so, and also because a general framework for thinking about perceptual models has been somewhat lacking (see below). We have done so to a limited extent. It is what led KB to abandon Darcy-Richards equation based models at an early age (see Beven, 1989b, 2001c), and to continue to work with Topmodel and Dynamic Topmodel, which can reflect some elements of the perceptual model described above in an integrated way. It is also why he has tried to convey the idea that those models are not applicable everywhere, despite attempts to incorporate them into global land surface parameterizations. NC by contrast has largely taken a more top-down approach of seeing what hydrological processes might be inferred from data-based mechanistic (DBM) models of dominant modes of observed response at the hillslope to catchment scales (e.g., Chappell et al., 2006, 2017; Ockenden & Chappell, 2011).

5 | CRITICAL QUESTIONS

It is clear that any definition of a perceptual model for a particular hydrological system at a certain scale will be complex, will include the recognition of nonlinear hydrological functioning, and will recognize the potential importance of

processes that are very difficult to observe and fluxes that are difficult to quantify accurately. Figure 2 essentially represents that complexity in a schematic way (with the additional element that the distributions and their linkages will change over time). On the other hand, we do have observations at certain scales (let us refer to these as a particular scales of "control volume") that can be analyzed, using either conceptual models or data-based methods, that aim to maximize the information extracted. But there is a missing link here. While these data-based methods (or other procedural models) capture some dominant modes of response of the catchments where applied, they do not necessarily tell us what are the essential processes of the bottom-up perceptual model. This is fundamentally linked to the question raised earlier of "getting the right results for the right reasons" in addition to any calibration/validation exercise (Beven, 1987; Kirchner, 2019; Klemeš, 1986, 1997; Oreskes et al., 1994), We would suggest that this perplexity then poses two critical questions that require addressing in future[1]:

1. How can perceptual model hypotheses be tested at scales of interest, or what constitutes consistency between hydrological observations and perceptual understanding of how the hydrological system functions?
2. How can constraints then be imposed on the basis of qualitative perceptual knowledge in conditioning predictive models?

We will consider these questions in turn.

6 | TESTING PERCEPTUAL MODEL HYPOTHESES

Consider the first question. There is an immediate problem. The perceptual model recognizes complexities at the small scale (which may still control the distribution of responses at larger scales, for example, Wood et al., 1988); the data analysis necessarily takes place at the scale of some larger control volume, commonly a catchment where there is a discharge gauging station (and sometimes a plot, or hillslope, or other experimental facility). Some of that complexity will be lost in the integration to the control volume scale but not necessarily in the sense of simple averaging. As in the concept of the Representative Elementary Area of Wood et al. (1988) the extremes of the small-scale variability might be important in controlling the overall response (as might be the case when preferential flows are important). This makes it difficult to use a perceptual model in a forward sense, that is, to incorporate the understanding of complexity into a quantitative model at the catchment scales. This is what has made it so difficult to demonstrate that distributed models based on "physical principles" can provide adequate predictions at the catchment scale (albeit that such models are often deficient in fully building in all the processes of the perceptual model, see for example, Beven, 2001a, 2001b, 2001c, 2006a, 2006b, 2018).

So, how to test if any element of a perceptual model is wrong? This should in fact be relatively simple in that any conflict between observed states or dynamics in a catchment and the perceptual model should result in an immediate (Popperian) falsification of that aspect of the perceptual model leading to its modification or removal. It is only our perception of how the catchment functions hydrologically that needs to change, accepting there is often considerable hydrological inertia in changing thought processes (Beven, 2018, 2020b, 2021). Of course, as with the example of the Horton infiltration theory above, this may not be apparent in the analysis or modeling of a discharge hydrograph. It will depend more on what is actually observable and whether we have the time and resources (or even inclination when there is a "pet model" to be calibrated) to make such observations. The occurrence of surface saturation or overland flow, for example, is in principle directly observable; much of what takes place in the subsurface is not and in both cases a few point observations might be quite misleading since they might not be in the right places or taken at the right times to properly capture the nature of the response. Further, we are often dependent on the more detailed observations of others in a catchment of interest (especially those who have actually been out in storms as they happen), or in catchments distant from the study area but with (ideally) climatologically, topographically or geologically "similar" characteristics.

What is much more difficult is to demonstrate what is *essential* in the perceptual model when we are interested in the responses dominating at the catchment scale. We are more often in the situation of making inferences about the perceptual model from the analysis of hydrological observations and modeling results, at the control volume scale of the whole catchment. That is not to say that we cannot evaluate consistency between our perceptions of hydrological functions and observations at that scale. Classical examples are recession curve analysis and flow duration curves or other hydrological signatures from observed discharges as indicators of storage change or effective volume in headwater catchments. This will, of course, be more informative where those signatures can be most easily be related to particular processes (via recession coefficients or time constants) and can be robust to uncertainties in the observations (Gupta

et al., 2008; Di Baldassarre & Montanari, 2009; Wagener et al., 2007; Westerberg & McMillan, 2015; Westerberg et al., 2016; McMillan, 2021).

We can also use the strategy of inference based on dynamic modeling. Approaches such as the data-based mechanistic (DBM) modeling of Young (2003), Young, 2013) can be used to determine appropriate model structures and associated parameter values for individual catchments. The recursive estimation techniques used are designed to be robust to uncertainties, and when used with time variable parameter identification can be used to explore the nonlinear gains in the system (e.g., Young & Beven, 1994, Chappell & Tych, 2012, and Mindham et al., 2018, for hydrological examples), while the time distribution is represented using linear transfer functions. Young (2011) has developed continuous time model identification algorithms for which the parameters are consistent across time steps. Model structures are evaluated in terms of both goodness-of-fit and uncertainty in the parameter estimates. Systems analysis of rainfall-runoff dynamics of catchment systems often result in a parallel pathways model structure with fast and slow time constants representing two different dominant modes of response, preceded by a nonlinear gain function. Such a structure requires only a small number of gain and time constant parameters that can inform a local perceptual model, or be used to classify catchments in terms of their modes of response. Such information is one top-down way of building up hydrological knowledge at integrated control volume scales in that it captures the dominant modes of response at that scale. It does not in itself, however, provide any inference about consistency with the detailed perceptual model of smaller scale processes. This would require more specific experimentation on those processes which could be used to feed information into the data-based modeling in a way that goes beyond the catchment characteristic indices (such as those used in the deep learning data-based model of Kratzert et al., 2019). We will consider this further in the next section.

By using recursive estimation as a way of studying an appropriate nonlinearity, and by allowing for parameter uncertainties, the DBM approach provides a neat way of identifying appropriate model structures for a particular system. In that the approach depends on time-series observations, it would be more difficult to apply to the relevant fluxes of energy and momentum, or to biogeochemical variables, since input and output sequences for hillslope and catchment scales are less readily available, and local observations (e.g., Fluxnet sites) may not be representative of larger areas. Other studies have used more traditional modeling toolkits. The study of Wrede et al. (2015), noted earlier, used the SuperFLEX modeling approach to identify model structures for three catchments in Luxembourg. This revealed a perceptual expectation that these catchments would require different model structures, primarily as a result of differences in their hydrogeology. This was borne out by the identified model structures. This study did, however, rely on the optimization of a goodness of fit index to identify appropriate model structures and associated parameter values, without any consideration of uncertainty in the calibration data used. This type of approach has been extended to automatic model structure identifications as well as parameter value optimization by Spieler et al. (2020). Knoben et al. (2020) used the MARRMoT ensemble of rainfall-runoff models in a study of 559 catchments selected from the US CAMELS database in this way, though they did compare the model results with a benchmark model of mean or modal daily values of the calibration data. In their case, model choice appeared to depend more on the choice of objective function than catchment characteristics. Lane et al. (2019) came to similar conclusions in applications of FUSE model structures to 1,000 UK catchments, also demonstrating that models might fail for catchments where there was poor water balance closure. Kratzert et al. (2019) suggest that a Deep Learning Model will out-perform any conceptual hydrological model across such an ensemble of catchment datasets. This might be, in part, because of the potential of data-based methods to compensate for consistent data errors (Beven, 2020a, 2020b, 2021).

These types of meta-analyses would be required to underpin the type of global perceptual model of expectations of dominant modes of response discussed in Wagener et al. (2021) who argue that model applications should explicitly state the type of perceptual understanding that underlies the choice of a model structure. In that way, the choice of a particular model structure should more properly reflect the perceptual model for a catchment. Fowler et al. (2020) point out that this may not always be the case, particularly in the case of representing the storages associated with long recession periods. This was also a result of applying a single semi-distributed model structure (DECIPHeR) to the CAMELS-GB 1000 catchment database in Coxon et al. (2019). Their study demonstrated the need to modify the model structure for catchments with large baseflow components.

7 | USING PERCEPTUAL UNDERSTANDING TO CONSTRAIN HYDROLOGICAL MODELS

The second critical question is how to use perceptual understanding to constrain the predictions and prediction uncertainties of hydrological models. In some circumstances, we are not only interested in the dominant modes of response

that can be determined from hydrological signatures or data-based models, but rather in predicting future behaviors as a result of management changes. That implies that there will be changes to processes at scales within the control volume that will then impact on the response. How the gains and time constants might change into the future is also an interesting question, though if we really thought about it, speculating about changes in such dominant mode parameters (Chappell et al., 2006; Jones et al., 2014; Walsh et al., 2011) might not be more difficult than speculating about changes in the many more interacting process parameters in a more reductionist approach (an argument made long ago by Beven, 1989b). We are just less used to thinking in a less reductionist way (see also the discussion in Wagener et al., 2021). Where process parameters in a model have been estimated by calibration, there might be complex interactions with the parameters of other process representations. It does remain a scientific challenge in hydrology, however, to show that we can model the processes of complex catchment systems (and potential future changes) successfully.

Additional observations more directly related to the perceptual model might then prove valuable in constraining hydrological model predictions in ways that "*show that we do, after all, understand our science and its complex interrelated phenomena*" (going back to Kohler, 1969). As Beven (1989b) suggested: "*I believe that any application in which physically correct predictions are considered important must involve a close cooperation between field observation and modeling. The nature of this cooperation, and in particular the nature of appropriate field observation and parameter measurement techniques, requires considerable further research if physically-based models are to realize the advantages that are currently being claimed for them.*" (p. 168). This implies using observations that go beyond input–output time series and static catchment indices and signatures (see, for example, Hingray et al., 2010; Euser et al., 2013; Shafii & Tolson, 2015) which can be problematic to relate to specific processes. This is related to suggestions of using "soft" data or "auxiliary" data in model calibrations and evaluations in the past (e.g., Mroczkowski et al., 1997; Seibert & McDonnell, 2002; Winsemius et al., 2009; Son & Sivapalan, 2007). Time series of internal state, flux or tracer measurements can then be related more directly to perceptual process understanding, even if the observations might be sparse.

This might be, for example, patterns of saturation and estimates of saturated areas (Ambroise et al., 1996; Beven & Kirkby, 1979; Blazkova et al., 2002; Lamb et al., 1998), the dynamics of discontinuous streams (Botter & Durighetto, 2020; Godsey & Kirchner, 2014; Van Meerveld et al., 2019); or the fraction of young water in the hydrograph for information about velocities as well as celerities (Von Freyberg et al., 2018). Techniques such as the ensemble hydrograph separation approach to tracer analysis of Kirchner (2019) can reveal something about time variability in responses that might provide new insights into perceptual models of transport. Knapp et al. (2019) used this approach with high frequency isotope data at Plynlimon and showed that only relatively small percentages of rainfalls over the previous 24 hr and week were contributing to streamflow in this wet, highly responsive catchment. Earlier isotope studies at Plynlimon on storm discharges in ephemeral natural peat pipes at Plynlimom had also shown a response dominated by displaced water (Sklash et al., 1986).

To use such information to constrain hydrological model predictions requires, however, that the model structures used can provide predictions of the required variables. In some cases this reflects the perceptual model used in defining a model, such as the saturated area predictions of Topmodel and Dynamic Topmodel, but those models that are based on collections of conceptual storages might not easily make such a link, that is, observed and modeled storage variables might not be directly commensurable.

So looking to the future, what types of information might be valuable in constraining hydrological models in ways consistent with perceptual understanding. The value of information in this way could be examined by a form of pre-posterior prior sensitivity analysis using a model structure set up to predict different types of variable that could potentially be observed in the future (see, e.g., Beven et al., 2020). This is somewhat different from more traditional analyses of sensitivity of modeled variables or objective functions to parameter variations (e.g., Kelleher et al., 2015; Maples et al., 2020; Pappenberger et al., 2006; Pianosi et al., 2015) and more like the assessment of where to drill a new well in aquifer modeling (Freeze et al., 1992; Rajabi et al., 2018). Such an analysis allows us to speculate, based on our perceptual understanding, about observations that could be worth testing in such a way. Such speculation suggests that local point measurements might be less valuable, unless they reflect the integration of processes around that point. Fluxnet or groundwater level observations do so to an extent but might not necessarily be representative in the cases of heterogeneous vegetation covers or shallow saturated systems. Very many point measurements might perhaps be valuable if they reveal some spatial coherence of responses, but this is not always the case even in a single catchment (e.g., the soil moisture profile sampling system of Western et al., 1999).

More integrative measures, such as mapped saturated areas and lengths of flowing streams, have been used in the past in model evaluations, but have generally relied on occasional ground-based surveys. It would be particularly valuable to have techniques for more continuous observations over larger spatial scales, such as using remote sensing to

detect surface water in some way. Both microwave sensing and infrared temperature sensing have been used in detecting surface soil moisture and groundwater return flows (e.g., Deng et al., 2020; Glaser et al., 2016) but the former is dependent on the apparent dielectric constant of the soil surface which shows a lack of sensitivity near saturation at MHz frequencies, and the latter is dependent on a consistent difference in temperature for water sources.

Another measure that could be valuable, not only in constraining models but also in closing the water balance, would be a direct measure of storage changes in a catchment. Gravity anomaly techniques have been used both on the ground (Güntner et al., 2017) and on the GRACE satellite. The latter has such a coarse resolution, however, that it has little value at small catchment scales. Ground-based methods have not been used widely, in part because it is a very expensive technique but also, perhaps, because there are more issues in getting time series of storage values. Testing the potential for this method in practice could be evaluated by a pre-posterior prior analysis but more widespread implementation would require much cheaper installation costs. There is now an increasing network of COSMOS soil moisture sensors that do give time series, but only for very near surface storage within a volume that changes with moisture status, COSMOS data have been used for model calibration (e.g., see the discussion of Dimitrova-Petrova et al., 2020) but there is a real issue of commensurability with predicted variables in this case. Microwave remote sensing of surface soil moisture suffers from similar issues of sensitivity and commensurability.

One technique that might prove cost effective could be the observation of incremental discharges in river networks. Even at small scales incremental discharges can show considerable variability, primarily as a result of hydrogeological variations along channels (e.g., Genereux et al., 1993). It therefore has the potential to both inform the perceptual model (do we really understand the differences that are observed?) and to provide constraints on predictive models. To take full advantage of the information, of course, the model would have to be set up in at least a semi-distributed way to make the observed and predicted reach discharges commensurable, and the discharges would have to be estimated with high accuracy (since incremental discharge requires taking differences, thereby increasing the uncertainty associated with the estimates).

The possibility of distributed predictions also brings in another possible source of perceptual information - knowledge from local stakeholders. The possibility of detailed visualizations of model outcomes gives rise to the possibility of those predictions being evaluated for consistency with local knowledge. In particular, conflicts between model predictions and local perceptual knowledge can lead to modifications to model processes, parameters, or boundary conditions (e.g., Lane et al., 2011). This type of falsification underlies the "models of everywhere" predictive modeling strategy advocated, for example, by Beven (2007), Beven and Alcock (2012) and Blair et al. (2019). It is worth making the point that even such simple evaluations might have avoided some perceptual errors of the past such as in the continuing application of infiltration excess models of streamflow generation in areas where widespread overland flow was simply not observed (e.g., discussion of Beven, 2021). The important factor in such a test, of course, is to actually *observe* what is going on during hydrologically active periods as a way of constraining model predictions and, in future, develop more techniques for making this possible.

8 | CONCLUSIONS

Hydrological models based on process representations currently face a challenge from data-based and machine learning modeling methodologies that will often provide a better fit to the available calibration and test data. There may be some good reasons for this, including the effects of noise and errors in observed hydrological data sets (see Beven, 2019, 2020a) but this challenge is compounded by the difficulties of subsuming the essential elements of highly complex perceptual models of the processes into a parsimonious predictive model. While we can easily modify the qualitative perceptual model of a catchment following direct observations, we do not have good ways of deciding what is *essential* to include in a predictive model; we do not have good ways of learning from data-based analyses of dominant modes of response to improve process-based models; and we do not have good ways of "learning from the physics" where that physics is small scale and dominated by unknowable local boundary conditions in heterogeneous subsurface and complex vegetated surfaces. We can however, use local conflicts between model simulations and perceptions of how a catchment works as a way of learning how to improve models of everywhere.

This is important in the context of predicting future change. So far, we have primarily considered perceptual models in the context of the physical processes of catchment response. Future change as a result of human activities whether directly or indirectly through climate shifts, requires additional components of a perceptual model of what the impact of such changes might be. This is the focus of the Panta Rhei initiative of IAHS (see, for example, Kreibich et al., 2017)

and part of the perceptual model uncertainty of Westerberg et al. (2017). While data-based and machine learning methods might well be able to demonstrate better performance in reproducing past hydrological observations, prediction of future change would require that enough information about the effects of such change is available in the training data set of past observations (see the discussion of Mount et al., 2016, in the context of Panta Rhei). Some such information may be available in a generalized way (proportions of urban or forested areas for example), but more specific spatially localized changes would not be (such as assessing different spatial strategies in natural flood management interventions). There is still a need therefore for process-based models that can reflect these localized forms of change more explicitly. The question is *how to implement a learning process in moving from perceptual complexity to parsimonious predictability?*

What can we say about that question now?

- We need better ways of observing spatial patterns of hydrological responses so as to refine our perceptual models and understanding of the dominant controls on those responses. Remote sensing might help here (e.g., Antonelli et al., 2020), but is often limited to only near-surface signals and subject to significant uncertainties (e.g., estimates of evapotranspiration) and limitations of the calibrations (e.g., sensitivity of the apparent dielectric constant to unknown variations in organic matter content). Other possibilities need to be considered, including low cost distributed sensors and more integrative measurements such as incremental stream discharges (e.g., Beven et al., 2020; Mao et al., 2019).
- There will be some processes, for example preferential flows in the subsurface, that will be particularly difficult to study and parameterize. Evidence for their importance is ubiquitous and they will feature in many hydrologists' perceptual models, not only in terms of water flows but also in the transport of nutrients and pollutants. Preferential flows can be generated by a variety of different mechanisms depending on soil structures and patterns of inputs, and their functioning is not readily quantified without disturbing the soil profile. It might therefore be necessary to subsume the detail of such processes as a parameterization at a larger "representative" scale (see, for example, Chappell et al., 1998; Beven, 2006b). It might be possible to infer the nature of that parameterization using data-based or machine learning methods at that scale (e.g., Beven, 2020a). Ways are needed to make novel types of observations and sufficient data available to allow such inference.
- We might be guided in what is essential, and in what type of novel observations might be useful, by the type of model sensitivity analysis discussed earlier. This, of course, comes with the important caveat that any such analysis is dependent on the model used and the assumptions that are necessarily incorporated when simplifying any underlying perceptual model.
- We should be prepared to accept that models, or particular parameter sets within a model, might be invalidated by being inconsistent with a perceptual model of how a catchment responds, even if they can be calibrated to give good fits to the streamflow data (there is actually an example of this back in Beven & Kirkby, 1979). As detailed distributed predictions become more common (the models of everywhere concept), invalidation will increasingly occur at a local level invoking changes to the model (or perhaps a closer look at the underlying perceptual model)

We can conclude that there remains much to be done in moving beyond traditional calibration of readily available model structures and testing only goodness of fit indices. We require new ways of incorporating perceptual understanding into parsimonious model structures as a way of moving closer to getting the right results for the right reasons.

CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS
Keith Beven: Conceptualization; writing-original draft; writing-review & editing. **Nick Chappell:** Conceptualization; project administration; writing-review & editing.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study

ORCID

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

Nick A. Chappell https://orcid.org/0000-0001-6683-951X

ENDNOTE

[1] Note that the term perceptual model does not appear anywhere in the Blöschl and Merz (2009) paper on 23 Unsolved Problems in Hydrology. A lack of perceptual understanding does appear to underlie a number of the problems listed, however, and the questions posed here are relevant particularly to Unsolved Problems Nos. 4, 5, 7, 8, 12, 13, 14, 15, 18, and 19.

[Correction added on 24 May 2021, after first online publication: Further Reading section has been removed.]

RELATED WIREs ARTICLES

On hypothesis testing in hydrology: Why falsification of models is still a really good idea

Hydrological data uncertainty and its implications

Historical development of rainfall-runoff modeling

REFERENCES

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*(11), 8792–8812.

Alley, W. M., Healy, R. W., LaBaugh, J. W., & Reilly, T. E. (2002). Flow and storage in groundwater systems. *Science*, *296*, 1985–1990.

Ahuja, L. R., & Ross, J. D. (1983). Effect of subsoil conductivity and thickness on interflow pathways, rates and source areas for chemicals in a sloping layered soil with seepage face. *Journal of Hydrology*, *64*(1–4), 189–204.

Ambroise, B., Freer, J., & Beven, K. J. (1996). Application of a generalised TOPMODEL to the small Ringelbach catchment, Vosges, France. *Water Resources Research*, *32*(7), 2147–2159.

Antonelli, M., Glaser, B., Teuling, A. J., Klaus, J., & Pfister, L. (2020). Saturated areas through the lens: 1 spatio-temporal variability of surface saturation documented through thermal infrared imagery. *Hydrological Processes*, *34*(6), 1310–1332.

Baker, V. R. (2017). Debates—Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty. *Water Resources Research*, *53*(3), 1770–1778.

Beven, K. J., (1987), Towards a new paradigm in hydrology. In: Water for the future: Hydrology in perspective, IAHS Publ. No. 164, 393–403.

Beven, K. J. (1989a). Interflow. In H. J. Morel-Seytoux (Ed.), *Proceedings of NATO ARW on unsaturated flow in hydrological modelling* (pp. 191–219). Dordrecht: Kluwer.

Beven, K. J. (1989b). Changing ideas in hydrology: The case of physically based models. *Journal of Hydrology*, *105*, 157–172.

Beven, K. J. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*, 41–51.

Beven, K. J. (2001a). *Rainfall-runoff modelling: The primer* (1st ed.). John Wiley & Son: Chichester.

Beven, K. J. (2001b). Calibration, validation and equifinality in hydrological modelling. In M. G. Anderson & P. D. Bates (Eds.), *Model validation: Perspectives in hydrological science* (pp. 43–55). Chichester: Wiley.

Beven, K. J. (2001c). Dalton medal lecture: How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*(1), 1–12.

Beven, K. J. (2004a). Robert Horton's perceptual model of infiltration. *Hydrological Processes*, *18*, 3447–3460.

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

Beven, K. J. (2006a). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*, 18–36.

Beven, K. J. (2006b). The holy grail of scientific hydrology: $Q_t = H(S\,R)A$ as closure. *Hydrology and Earth Systems Science*, *10*, 609–618.

Beven, K. J. (2007). Working towards integrated environmental models of everywhere: Uncertainty, data, and modelling as a learning process. *Hydrology and Earth System Science*, *11*(1), 460–467.

Beven, K. J. (2010). Preferential flows and travel time distributions: Defining adequate hypothesis tests for hydrological process models. *Hydrological Processes*, *24*, 1537–1547.

Beven, K. J. (2015). What we see now: Event-persistence in predicting the responses of hydro-eco-geomorphological systems? *Ecological Modelling*, *298*, 4–15.

Beven, K. J. (2018). A century of denial: Preferential and nonequilibrium water flow in soils, 1864–1984. *Vadose Zone Journal*, *17*, 180153. https://doi.org/10.2136/vzj2018.08.015

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

Beven, K. J. (2020a). Deep learning, hydrological processes and the uniqueness of place. *Hydrological Processes*, *34*, 3608–3613. https://doi.org/10.1002/hyp.13805

Beven, K. J. (2020b). A history of the concept of time of concentration. *Hydrology and Earth System Sciences*, *24*, 2655–2670. https://doi.org/10.5194/hess-24-2655-2020

Beven, K. J. (2021). The era of infiltration. *Hydrology and Earth System Sciences*, *25*(2), 851–866. https://doi.org/10.5194/hess-2020-308

Beven, K. J., & Alcock, R. (2012). Modelling everything everywhere: A new approach to decision making for water management under uncertainty. *Freshwater Biology*, *56*, 124–132. https://doi.org/10.1111/j.1365-2427.2011.02592.x

Beven, K. J., Asadullah, A., Bates, P. D., Blyth, E., Chappell, N. A., Child, S., ... Wagener, T. (2020). Developing observational methods to drive future hydrological science: Can we make a start as a community? *Hydrological Processes*, *34*, 868–873. https://doi.org/10.1002/hyp.13622

Beven, K. J., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, *J. Hydrology*, *249*, 11–29.

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., & Kirkby, M. J. (1979). A physically-based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69.

Beven, K. J., & Smith, P. J. (2015). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *ASCE's Journal of Hydrologic Engineering*, *20*(1), A4014010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991

Black, P. E. (1997). Watershed functions. *Journal of the American Water Resources Association*, *33*(10), 1–11.

Blair, G. S., Beven, K. J., Lamb, R., Bassett, R., Cauwenberghs, K., Hankin, B., ... Towe, R. (2019). Models of everywhere revisited: A technological perspective. *Environmental Modelling and Software*, *122*, 104521. https://doi.org/10.1016/j.envsoft.2019.104521

Blazkova, S., Beven, K., Tacheci, P., & Kulasova, A. (2002). Testing the distributed water table predictions of TOPMODEL (allowing for uncertainty in model calibration): The death of TOPMODEL? *Water Resources Research*, *38*(11), W01257. https://doi.org/10.1029/2001WR000912

Blöschl, G., & Merz, R. (2009). Landform–hydrology feedbacks. In J.-C. Otto & R. Dickau (Eds.), *Landform-structure, evolution, process control* (pp. 117–126). Berlin, Heidelberg: Springer.

Bonell, M., & Williams, J. (1986). The generation and redistribution of overland flow on a massive oxic soil in a eucalypt woodland within the semi-arid tropics of North Australia. *Hydrological Processes*, *1*(1), 31–46.

Botter, G., & Durighetto, N. (2020). The stream length duration curve: A tool for characterizing the time variability of the flowing stream length. *Water Resources Research*, *56*(8), e2020WR027282. https://doi.org/10.1029/2020WR027282

Brammer, D. D., & McDonnell, J. J. (1996). An evolving perceptual model of hillslope flow at the Maimai catchment. *Advances in Hillslope Processes*, *1*, 35–60.

Buytaert, W., & Beven, K. J. (2009). Regionalisation as a learning process. *Water Resources Research*, *45*, W11419. https://doi.org/10.1029/2008WR007359

Chappell, N. A. (2010). Soil pipe distribution and hydrological functioning within the humid tropics: A synthesis. *Hydrological Processes*, *24*(12), 1567–1581.

Chappell, N. A., Franks, S. W., & Larenus, J. (1998). Multi-scale permeability estimation in a tropical catchment. *Hydrological Processes*, *12*, 1507–1523.

Chappell, N. A., & Lancaster, J. W. (2007). Comparison of methodological uncertainty within permeability measurements. *Hydrological Processes*, *21*(18), 2504–2514.

Chappell, N. A., & Tych, W. (2012). Identifying step changes in single streamflow and evaporation records due to forest cover change. *Hydrological Processes*, *26*, 100–116. https://doi.org/10.1002/hyp.8115

Chappell, N. A., Tych, W., Chotai, A., Bidin, K., Sinun, W., & Thang, H. C. (2006). BARUMODEL: Combined data based mechanistic models of runoff responsein a managed rainforest catchment. *Forest Ecology and Management*, *224*, 58–80.

Chappell, N. A., Jones, T. D., Tych, W., & Krishnaswamy, J. (2017). Role of rainstorm intensity underestimated by data-derived flood models: Emerging global evidence from subsurface-dominated watersheds. *Environmental Modelling and Software*, *88*, 1–9. https://doi.org/10.1016/j.envsoft.2016.10.009

Chorley, R. J. (1978). The Hillslope Hydrological Cycle. In M. J. Kirkby (Ed.) *Hillslope Hydrology* Chichester: John Wiley & Sons, 1–42.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Arnold, J. R. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498–2514.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, *51*(4), 2515–2542.

Cook, H. L. (1946). The infiltration approach to the calculation of surface runoff. *Transactions of the American Geophysical Union*, *27*, 726–743.

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., ... Woods, R. (2019). DECIPHeR v1: Dynamic fluxEs and Connectivity for predictions of HydRology. *Geoscientific Model Development*, *12*, 2285–2306. https://doi.org/10.5194/gmd-12-2285-2019

Deng, Y., Wang, S., Bai, X., Wu, L., Cao, Y., Li, H., ... Tian, S. (2020). Comparison of soil moisture products from microwave remote sensing, land model, and reanalysis using global ground observations. *Hydrological Processes*, *34*(3), 836–851.

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*, 913–921.

Dimitrova-Petrova, K., Geris, J., Wilkinson, E. M., Rosolem, R., Verrot, L., Lilly, A., & Soulsby, C. (2020). Opportunities and challenges in using catchment-scale storage estimates from cosmic ray neutron sensors for rainfall-runoff modelling. *Journal of Hydrology*, *586*, 124878. https://doi.org/10.1016/j.jhydrol.2020.124878

Dixon, R. M., & Linden, D. R. (1972). Soil air pressure and water infiltration under border irrigation. *Soil Science Society of America Journal*, *36*(6), 948–953.

Dosher, B., & Lu, Z. L. (2017). Visual perceptual learning and models. *Annual Review of Vision Science*, *3*, 343–363.

Dunne, T., & Black, R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*, *6* (5), 1296–1311.

Dunne, T., Zhang, W., & Aubry, B. F. (1991). Effects of rainfall, vegetation, and microtopography on infiltration and runoff. *Water Resources Research*, *27*(9), 2271–2285.

Eagleson, P. S. (1970). *Dynamic hydrology*. New York, NY: McGraw-Hill.

Emmett, W. W. (1978). Overland flow. In M. J. Kirkby (Ed.), *Hillslope hydrology* (pp. 145–176). Chichester: John Wiley and Sons.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, *17*(5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013

Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L., & Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, *52*(3), 1820–1846.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., ... Zhang, L. (2018). Simulating runoff under changing climatic conditions: A framework for model improvement. *Water Resources Research*, *54*(12), 9812–9832.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., ... Western, A. (2020). Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections. *Water Resources Research*, *56*(5), e2019WR025286

Freer, J., Beven, K. J., & Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resources Research*, *32*(7), 2161–2173.

Freeze, R. A., James, B., Massmann, J., Sperling, T., & Smith, L. (1992). Hydrogeological decision analysis: 4. The concept of data worth and its use in the development of site investigation strategies. *Groundwater*, *30*(4), 574–588.

Genereux, D. P., Hemond, H. F., & Mulholland, P. J. (1993). Spatial and temporal variability in streamflow generation on the west branch of the Walker Branch Watershed. *J. Hydrology*, *142*, 137–166.

Gilman, K., & Newson, M. D. (1980). *Soil pipes and pipeflow—A hydrological study in Upland Wales. British Geomorphological Research Group, Research Monograph 1*. Norwich: Geobooks.

Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L., & Hopp, L. (2016). On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. *Water Resources Research*, *52*(10), 8317–8342.

Gnann, S. J., McMillan, H. K., Woods, R. A., & Howden, N. J. (2021). Including regional knowledge improves baseflow signature predictions in large sample hydrology. *Water Resources Research*, *57*(2), e2020WR028354. https://doi.org/10.1029/2020WR028354

Godsey, S. E., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*, 5791–5803. https://doi.org/10.1002/hyp.10310

Güntner, A., Reich, M., Mikolaj, M., Creutzfeldt, B., Schroeder, S., & Wziontek, H. (2017). Landscape-scale water balance monitoring with an iGrav superconducting gravimeter in a field enclosure. *Hydrology and Earth System Sciences*, *21*(6), 3167.

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, *52*(3), 218–242.

Hewlett, J. D., & Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. In W. E. Sopper & H. W. Lull (Eds.), *Forest hydrology* (pp. 275–290). Oxford: Pergamon Press.

Hingray, B., Schaefli, B., Mezghani, A., & Hamdi, Y. (2010). Signature-based model calibration for hydrological prediction in mesoscale alpine catchments. *Hydrological Sciences Journal*, *55*(6), 1002–1016.

Horton, R. E. (1942). Remarks on hydrologic terminology. *Transactions of the American Geophysical Union*, *23*, 479–482.

Horton, R. E., Leach, H. R., & van Vliet, R. (1934). Laminar sheet flow. *Transactions of the American Geophysical Union*, *15*, 393–404.

Johnson, M. S., & Lehmann, J. (2006). Double-funneling of trees: Stemflow and root-induced preferential flow. *Ecoscience*, *13*(3), 324–333. https://doi.org/10.2980/i1195-6860-13-3-324.1

Jones, J. A. A. (2010). Soil piping and catchment response. *Hydrological Processes*, *24*(12), 1548–1566.

Jones, T. D., Chappell, N. A., & Tych, W. (2014). First dynamic model of dissolved organic carbon derived directly from high frequency observations through contiguous storms. *Environmental Science & Technology*, *48*, 13289–13297. https://doi.org/10.1021/es503506m

Käser, D. H., Binley, A., Krause, S., & Heathwaite, A. L. (2014). Prospective modelling of 3D hyporheic exchange based on high-resolution topography and stream elevation. *Hydrological Processes*, *28*(4), 2579–2594.

Kelleher, C., Wagener, T., & McGlynn, B. (2015). Model-based analysis of the influence of catchment properties on hydrologic partitioning across five mountain headwater sub-catchments. *Water Resources Research*, *51*, 51–4136. https://doi.org/10.1002/2014WR016147

Kirchner, J. W. (2019). Quantifying new water fractions and transit time distributions using ensemble hydrograph separation: Theory and benchmark tests. *Hydrology and Earth System Sciences*, *23*(1), 303–349.

Klemeš, V. (1986). Dilettantism in hydrology: Transition or Destiny. *Water Resources Research*, *22*(9), 177S–188S.

Klemeš, V. (1997). Of carts and horses in hydrologic Modelling. *Journal of Hydraulic Engineering ASCE*, *2*(2), 43–49.

Knapp, J. L., Neal, C., Schlumpf, A., Neal, M., & Kirchner, J. W. (2019). New water fractions and transit time distributions at Plynlimon, Wales, estimated from stable water isotopes in precipitation and streamflow. *Hydrology and Earth System Sciences*, *23*(10), 4367–4388.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*(9), e2019WR025975. https://doi.org/10.1029/2019WR025975

Kohler, M. A. (1969). Keynote Address in Hydrological Forecasting. *World Meteorological Organisation Technical Note No. 92*, XI–XVI.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110.

Kreibich, H., Krueger, T., Van Loon, A., Mejia, A., Liu, J., McMillan, H., & Castellarin, A. (2017). Scientific debate of Panta Rhei research–how to advance our knowledge of changes in hydrology and society? *Hydrological Sciences Journal*, *62*(3), 331–332.

Lamb, R., Beven, K. J., & Myrabø, S. (1998). Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources*, *22*(4), 305–317.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., ... Reaney, S. M. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences*, *23*, 4011–4032. https://doi.org/10.5194/hess-23-4011-2019

Lane, S. N., Odoni, N., Landström, C., Whatmore, S. J., Ward, N., & Bradley, S. (2011). Doing flood risk science differently: An experiment in radical scientific method. *Transactions of the Institute of British Geographers*, *36*(1), 15–36.

Mao, F., Kieran Khamis, K., Krause, S., Clark, J., & Hannah, D. M. (2019). Low-cost environmental sensor networks: Recent Advances and Future Directions. *Frontiers in Earth Science*. *7*, 221. https://doi.org/10.3389/feart.2019.00221

Maples, S., Foglia, L., Fogg, G. E., & Maxwell, R. M. (2020). Sensitivity of hydrologic and geologic parameters on recharge processes in a highly-heterogeneous, semi-confined aquifer system. *Hydrogeology and Earth Systems Sciences*, *24*, 2437–2456. https://doi.org/10.5194/hess-24-2437-2020

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

McGlynn, B. L., McDonnell, J. J., & Brammer, D. D. (2002). A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments. *New Zealand Journal of Hydrology*, *257*(1–4), 1–26.

McMillan, H. K. (2021). A review of hydrologic signatures and their applications. *Wiley Interdisciplinary Reviews: Water*, *8*(1), p. e1499. https://doi.org/10.1002/wat2.1499

Mindham, D. A., Tych, W., & Chappell, N. A. (2018). Extended state dependent parameter modelling with a data-based mechanistic approach to non-linear model structure identification. *Environmental Modelling and Software*, *104*, 81–93.

Mosley, M. P. (1982). Subsurface flow velocities through selected forest soils south Island. *New Zealand Journal of Hydrology*, *55*, 65–92.

Mount, N. J., Maier, H. R., Toth, E., Elshorbagy, A., Solomatine, D., Chang, F. J., & Abrahart, R. J. (2016). Data-driven modelling approaches for socio-hydrology: Opportunities and challenges within the Panta Rhei science plan. *Hydrological Sciences Journal*, *61*(7), 1192–1208.

Mroczkowski, M., Raper, G. P., & Kuczera, G. (1997). The quest for more powerful validation of conceptual catchment models. *Water Resources Research*, *33*, 2325–2335.

Ockenden, M. C., & Chappell, N. A. (2011). Identification of the dominant runoff pathways from the data-based mechanistic modelling of nested catchments in temperate UK. *Journal of Hydrology*, *402*, 71–79. https://doi.org/10.1016/j.jhydrol.2011.03.001

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646.

Page, T., Chappell, N. A., Beven, K. J., Hankin, B., & Kretzschmar, A. (2020). Assessing the significance of wet-canopy evaporation from forests during extreme rainfall events for flood mitigation in mountainous regions of the UK. *Hydrology and Earth System Sciences*, *34*(24), 4740–4754. https://doi.org/10.1002/hyp.13895

Pappenberger, F., Iorgulescu, I., & Beven, K. J. (2006). Sensitivity analysis based on regional splits (SARS) and regression trees. *Environmental Modelling and Software*, *21*, 976–990.

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, *70*, 80–85.

Rahman, M., & Rosolem, R. (2017). Towards a simple representation of chalk hydrology in land surface modelling. *Hydrology and Earth System Sciences*, *21*, 459–471.

Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2018). Model-data interaction in groundwater studies: Review of methods, applications and future directions. *Journal of Hydrology*, *567*, 457–477.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, *15*, 2895–2911. https://doi.org/10.5194/hess-15-2895-2011

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), 1241. https://doi.org/10.1029/2001WR000978

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, *51*(5), 3796–3814.

Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*, 2101–2111.

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

Son, K., & Sivapalan, M. (2007). Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. *Water Resources Research*, *43*(1), W01415. https://doi.org/10.1029/2006WR005032

Spieler, D., Mai, J., Craig, J. R., Tolson, B. A., & Schütze, N. (2020). Automatic model structure identification for conceptual hydrologic models. *Water Resources Research*, *56*(9), e2019WR027009. https://doi.org/10.1029/2019WR027009

Tischendorf, W. G. (1969). Tracing storm flow to varying source area in small forested watershed in the Southeastern Piedmont. *Unpublished PhD dissertation*, Athens, GA: University of Georgia.

Tischer, A., Michalzik, B., & Lotze, R. (2020). Nonuniform but highly preferential stemflow routing along bark surfaces and actual smaller infiltration areas than previously assumed: A case study on European beech (*Fagus sylvatica* L.) and sycamore maple (*Acer pseudoplatanus* L.). *Ecohydrology*, *13*(6), e2230.

Vaché, K. B., & McDonnell, J. J. (2006). A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research*, *42*(2), W02409. https://doi.org/10.1029/2005WR004247

Van Meerveld, H. J. I., Kirchner, J. W., Vis, M. J., Assendelft, R. S., & Seibert, J. (2019). Expansion and contraction of the flowing stream network alter hillslope flowpath lengths and the shape of the travel time distribution. *Hydrology and Earth System Sciences*, *23*(11), 4825–4834.

Von Freyberg, J., Allen, S. T., Seeger, S., Weiler, M., & Kirchner, J. W. (2018). Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments. *Hydrology and Earth System Sciences*, *22*(7), 3841–3861.

Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N., Pianosi, F., ... Woods, R. (2021). On doing hydrology with dragons: Realizing the value of perceptual models and knowledge accumulation. *WIRES Water* in press.

Walsh, R. P. D., Bidin, K., Blake, W. H., Chappell, N. A., Clarke, M. A., Douglas, I., ... Annammala, K. V. (2011). Longer-term responses of rainforest erosional systems at different spatial scales to selective logging and climatic change. *Philosophical Transactions of the Royal Society B: Biological Science*, *366*, 3340–3353. https://doi.org/10.1098/rstb.2011.0054

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, *52*(3), 1847–1865.

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.

Western, A. W., Grayson, R. B., Blöschl, G., Willgoose, G., & McMahon, T. A. (1999). Observed spatial organisation of soil moisture and its relation to terrain indices. *Water Resources Research*, *35*, 797–810.

Winsemius, H. C., Schaefli, B., Montanari, A., & Savenije, H. H. G. (2009). On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research*, *45*(12), W12422. https://doi.org/10.1029/2009WR007706

WMO. (1974), International glossary of hydrology, WMO Report No. 385, Geneva.

Wood, E. F., Sivapalan, M., Beven, K. J., & Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modelling. *Journal of Hydrology*, *102*, 29–47.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., ... Pfister, L. (2015). Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. *Hydrological Processes*, *29*, 2731–2750.

Young, P. C. (2003). Top-down and data-based mechanistic modelling of rainfall–flow dynamics at the catchment scale. *Hydrological Processes*, *17*(11), 2195–2217.

Young, P. C. (2011). Recursive estimation and time-series analysis: An introduction for the student and practitioner. Heidelberg: Springer.

Young, P. C. (2013). Hypothetico-inductive data-based mechanistic modeling of hydrological systems. *Water Resources Research*, *49*(2), 915–935.

Young, P. C., & Beven, K. J. (1994). Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics*, *5*(3), 335–363.

Zaslavsky, D., & Sinai, G. (1981). Surface hydrology (5 parts). *Journal of Hydraulics Division, ASCE*, *107*, 1–93.