INVITED COMMENTARY

HYDROLOGICAL PROCESSES
Hydrol. Process. **17**, 1869–1875 (2003)
DOI: 10.1002/hyp.5132

Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response

Jeffrey J. McDonnell*

Department of Forest Engineering, Oregon State University, Corvallis, OR, USA

*Correspondence to:
Jeffrey J. McDonnell, Department of Forest Engineering, Oregon State University, Corvallis, Oregon 97330, USA. E-mail: Jeff.McDonnell@orst.edu

Introduction

Where does water go when it rains? What flow path does it take to the stream? How long does it reside in the catchment? These questions were articulated by John Hewlett within the context of his variable source area (VSA) concept almost 40 years ago (Hewlett and Hibbert, 1967). Today, we still grapple with these often vexing questions—now using new tools and approaches, but, as then, still searching for answers. Rapid progress is being made on the rainfall-runoff modelling front in catchment hydrology *vis-à-vis* parameter estimation techniques, model uncertainty analysis, examination of parameter identifiability in our models, downward approaches to hydrologic prediction, etc. (Beven 2001; Sivapalan, 2003). However, I wonder if we have somewhat neglected updating our understanding of the rainfall-runoff process and how this informs our needed model structures and response to these three basic questions central to our conceptualization of how catchments work?

One could argue that our sharpening perception of water source, flowpath and age in upland headwater catchments is radically different to what the framers of the VSA theory thought a half century ago (i.e. Hewlett in the USA, Cappus in France and Tsukamoto in Japan). Our best models still rely on mechanistic notions underlying the VSA, including saturation excess overland flow and subsurface stormflow (I will avoid mentioning how our operational models often are based on another whole older generation of streamflow generation concepts related exclusively to Horton!). The VSA concept has been distilled into our widely used research model structures by collapsing the process complexity into simple mathematical assumptions of things like the decline in saturated hydraulic conductivity with depth, steady-state catchment water table response, topographically defined water flowpaths and linear wetting and drying from the valley bottom upwards to the ridge (depending upon storm size, intensity and antecedent wetness conditions).

Much discussion is now devoted in the modelling literature towards the balance between practical simplifications of the VSA details and justifiable model complexity. This commentary takes a critical look at our process underpinning by discussing new field evidence of where water goes when it rains that directly challenges the *status quo*. New model structures informed by this new process understanding are then discussed in the context of how data

Accepted 14 April 2003

and objective discretization of catchment units may be used both to structure and test the model. The reader should note that this commentary is written by an experimentalist clearly biased towards headwater upland humid environments (but nevertheless like those watersheds for which VSA theory was first developed in Europe, Japan and the USA).

New Field Evidence of Where Water Goes When It Rains?

From the perspective of an experimentalist, the model structure often appears to be the most *ad hoc* part of the rainfall-runoff model exercise. The past decade has been dominated by a one-model-structure-fits-all view (e.g. TOPMODEL), with that global structure often being that of the VSA concept. Experimentalists have partly been to blame for this. It seems that in the intervening period between the advent of the VSA in the 1960s and now, we have bombarded the literature with all manner of interesting and ever-more detailed studies of complex, site-specific processes, making distillation of an emerging post-VSA concept difficult. Our mantra has been to document the idiosyncrasies of *yet another* experimental catchment (and then to report this in a paper), rather than seek commonality of response among the watersheds that we investigate through catchment intercomparison. The poor modeller is often forced to resort to International Hydrologic Decade (IHD)-era process representations (e.g. those reviewed in Dunne (1978)) because at least they were clear, unambiguous and compelling!

New field evidence of water source, flowpath and age suggests something quite different to the benchmark notions of runoff generation via mechanisms elucidated during the IHD. Perhaps the greatest paradigm shift has been the recognition that pre-event water largely dominates storm runoff. This changes everything, as Kirchner (2003) notes, whereby catchments store water for considerable periods of time but then release it promptly during storm events. Again, the benchmark studies that gave us the mechanisms (now well entrenched in our literature) of infiltration excess overland flow, saturation excess overland flow and subsurface stormflow (under the broad umbrella of the VSA concept) by pioneers in the 1960s and 1970s, like Betson, Dunne and Weyman, do not help us in our quest to explain this paradox of prompt release of old water within a flashy hydrograph. The hitherto avoidance of a process description in the z direction (i.e. any depth and hence volume component in the classical 'cartoon' representations of runoff generation processes in many of our textbooks) is a fundamental limitation to their use in defining today's model structures.

Below, I organize what I argue are some new ways of viewing the catchment rainfall-runoff processes by mapping this against some of the more prominent process assumptions in our VSA-structured models. These process assumptions include the way that modelers have encapsulated VSA theory into their numerics: that near-stream water table rises to intersect the soil surface with corresponding filling of storage and catchment-wide water table increases (what I call the catchment-wide steady state assumption), how near-stream inputs are then augmented by hillslope inputs, where hillslope flowpaths follow topography downslope (the topographic index assumption), that hillslope contributions "grow" from the near-stream zone upslope (the saturated wedge assumption) and that all of the processes of interest occur on the soil surface or in the soil mantle (the soil exclusivity assumption). The reader will see an emerging view of hillslopes and watersheds as non-linear systems. A critical post-VSA assessment of the rainfall-runoff process shows that streamflow outputs are not proportional to the inputs across the entire range of outputs. In other words, recent work is suggesting that watersheds display considerable threshold behaviour, storage effects, competitive feedbacks and hysteresis, in ways very much described by non-linearity theory (Phillips, 2003).

Steady-state catchment water table assumption

In most VSA-based conceptual runoff models, an unambiguous, monotonic function between the groundwater storage and runoff is implemented. Consequently, the dynamics of the simulated runoff from the groundwater zone always follows

the simulated rise and fall in groundwater levels. Increasingly, field evidence is challenging this notion. Seibert *et al.* (2003) showed that water table response in the riparian zone is often separate and independent from those positions farther upslope. We see this chemically as well. More than a few recent studies reporting end-member mixing results (Burns *et al.*, 2001) have shown that hillslope waters are chemically and isotopically distinct from riparian zone waters and that the degree of expression of hillslope water in the stream is minimal, or varies along a riparian aquifer volume gradient from watershed to watershed. Flux from the riparian zone often leads the hydrograph in a hysteretic way, with the hillslope input (on those storms when it is activated) dominating the recession limb after the threshold for its activation is exceeded (McGlynn and McDonnell, 2003).

Topographic index assumption

The widespread availability of digital elevation models (DEMs) has led many to use topography as a surrogate for water flow paths. In upland terrain this is reasonable theoretically, because elevation potential largely dominates total potential. But, recent work at Tarrawarra in Australia (http://www.civag.unimelb.edu.au/∼western/tarrawarra/tarrawarra.html) has led a number of researchers to question the notion that surface topography explains soil moisture distribution, let alone mobile lateral flow. At trenched hillslopes at places like Maimai in New Zealand, Panola in the USA, Fudoji in Japan, UBC School Forest and Plastic Lake in Canada, the bedrock topography, and not the surface topography, seems to be the most important surface for controlling the routing of mobile water laterally downslope. This is because transient water tables at these sites develop at the soil–bedrock interface and this transient saturation (i.e. the mobile lateral flow) then follows the microtopographic relief of the underlying surface laterally downslope (Freer *et al.*, 2002). These waters are often isotopically old (at least looking like water stored in the hillslope prior to the rainfall event), with waters emanating from those bedrock flowpaths often chemically dilute compared with the seepage from neighbouring zones due to their frequency of flushing (Burns *et al.*, 1999).

Saturated wedge assumption

There are several reasons to expect the development and growth of saturated wedges in the lower footslopes as defined in some of the benchmark UK studies of the IHD. While this certainly makes sense in and around the riparian zone, increasing evidence points to the hillslope in much less of a Weyman-*esque* saturated wedge-like fashion—one much more controlled by threshold lateral matrix and pipeflow behaviour at the hillslope scale. Indeed, when hillslopes ‘turn on’, many would argue that the rapid lateral flow is highly threshold dependent and largely via soil pipes located within the transient saturated zone or via discontinuities at the soil–bedrock interface (Uchida *et al.*, 2001). Admittedly, much of our trenching work in the past decades has yielded puzzling and equivocal results along these lines. Recent work reporting a sufficiently large number of storms to detect temporal patterns in hillslope response (Tromp van Meerveld and McDonnell, submitted) has shown clear and unambiguous thresholds of precipitation amount necessary to activate lateral flow on hillslopes.

Soil exclusivity assumption

The VSA concept, and the models derived from it, generally assumes that all processes relevant to rainfall-runoff occur within the soil mantle. The assumption of exponentially decreasing soil hydraulic conductivity with depth has been a way to implement this numerically. Soil hydraulic conductivity certainly does decline with depth in many upland soil environments (as determined *in situ*). However, when soil overlies bedrock that is not fully impermeable (the norm, I would argue, rather than the exception!), then our mean residence time studies often suggest that stream baseflow is considerably ‘older’ than would be possible by the soil mantle volume only. This means that how water below the soil profile connects vertically and laterally matters greatly to what we see in the channel in many environments (see recent work in the USA by Montgomery *et al.* (1997) and

in Japan by Onda *et al.* (2001)). I would argue that these observations defy explanation via the soil exclusivity assumption and VSA theory.

Moving towards a view of the watershed as a series of cryptic reservoirs

The plethora of experimental studies since the IHD have produced little generalizable potential for a post-VSA view of how watersheds work or even a definition of what appropriate state variables may be applicable in different environments. Worse, perhaps, is that we experimentalists have not yet articulated what are the minimal sets of measurements necessary to even characterize a single hillslope! Despite numerous calls by Tom Dunne (beginning with Dunne (1983)) and others for the past two decades, the dialogue between experimentalist and modeller still appears out of reach. Little has yet been done to merge experimental and modelling approaches. The experimentalist often proposes a perceptual model based on his or her complex and qualitative field observations and experiences, but the modeller usually does not incorporate the experimentalist's knowledge into the model structure, let alone the model calibration or testing. I try to summarize some possible ways forward below.

Flexible box models

I suggest that we need a formal replacement of VSA theory, perhaps moving towards a view of the catchment as a series of cryptic reservoirs that have coupled unsaturated and saturated zones, explicit dimensions and porosities, and that connect vertically and laterally in time and space in linear and non-linear ways. This seems to be a way to capture the first-order controls on what we observe from our experimental work, constrained by physical, chemical and isotopic data. This also allows explicit water and tracer mass balance within the model, making it just as physically based as more complex Darcy–Richards equation schemes. I would argue that whilst box models have been around for decades, their development and use in engineering-based watershed studies in the past was more out of algorithmic convenience rather than an enlightened process-mechanistic understanding. Notwithstanding, a box, tank, bucket or reservoir may be just the way forward to match the appropriate level of understanding and behaviour of our systems. I would argue that a box, objectively defined by distinct groundwater dynamics, soil solution chemistry or isotopic composition, with defined area, depth and porosity is a much better model building block than a multitude of elements over landscapes that are notoriously heterogeneous both vertically and laterally! Hydrograph recession analysis could be another way to define model boxes objectively in a downward way (see Sivapalan (2003)) and definition of the integrated characteristics of storage dynamics and hydrogeological features of aquifers in a watershed. Although most recession work has focused on mathematical 'fits' of the recession limb, Vitvar *et al.* (2002) has shown that the recession characteristics may contribute to the identification of distinct storage volumes and their control on expressed mean residence times for water in the channel.

An example of how we might consider a flexible model structure (the opposite to a one-size-fits-all VSA structure) was shown recently for the Maimai catchment by Seibert and McDonnell (2002). Earlier process work at the site had shown that riparian zone water table response and hillslope water table response were qualitatively different. Sampling of soil solution chemistry suggested, furthermore, that planar hillslopes, geomorphic hollows and riparian zones had a different chemistry, as well as (statistically significant) different classes of water isotopic composition. This constrained way of discretizing the catchment into reservoirs allowed for the construction of a reservoir-based model where a hillslope box filled and spilled laterally to a hollow box that, in turn, cascaded into a riparian zone box and then the stream. Based on the dialogue between modeller and experimentalist in this instance, Seibert and McDonnell (2002) were able to reduce the dimensionality of the problem to those discrete units that connected and disconnected in linear and non-linear ways. Although appropriate for the Maimai watershed, this structure would be most inappropriate for, say, the Sleepers River Watershed in the USA or the Rietholzbach watershed in Switzerland, where, unlike the Maimai, the bedrock is permeable and other assemblages of boxes would make sense to represent the age, as well as the source of flow in the channel.

Ultimately, the number of boxes one chooses is kept to a minimum if one adheres to a strict notion of predictive and model uncertainty. The idea though is that the structure *can* mimic the perception of the system, if this is constrained by physical, chemical and isotopic information. Modellers have shown us quite clearly that a model calibrated on discharge data alone is a weak test of whether the model is working correctly (Kirchner *et al.*, 1996). I would argue that experimentalists who claim 'understanding' of the watershed based on purely physical data (e.g. the relation between groundwater response and streamflow), or purely chemical data (e.g. an end member mixing analysis of geographic source), or on purely isotope data (e.g. a hydrograph separation into time sources or mean residence time analysis) are likewise fooling themselves—if we have learned anything in the past decade it is that constraining our perceptual model with flow, source and age together is what is needed for a robust process description of watershed function.

Soft ('fuzzy') data

One reason why, perhaps, the VSA concept and its implementation in many of our most popular research models has been with us so long is that calibration of a model on only the discharge signal rarely challenges the model structure (that we are often quite wedded to!) directly. Increasingly, research groups are starting to use fuzzy and uncertain information from field campaigns (like mapped near-stream saturated area, groundwater information, stream tracer information) as soft data (as opposed what one might call hard data in the form of a continuous streamflow signal). Seibert and McDonnell (2002) argue that soft data provide additional measures of model evaluation and parameter value acceptability beyond relative error measures such as the Nash and Sutcliffe (1970) efficiency and correlation coefficients. For instance, a high efficiency may be produced when calibrating a model only on runoff; however, by incorporating fuzzy data criteria in the model calibration, a better overall performance may be achieved, as interpreted by the experimentalist's view of runoff processes.

For instance, Seibert and McDonnell (2002) showed that seemingly good fits with model parameters optimized with only runoff (i.e. hard data) showed, in general, poor goodness of fit measures for other criteria such as the simulated event water contributions to peak runoff. Inclusion of soft data criteria in the model calibration process resulted in slightly lower overall efficiencies, but accepting lower efficiencies for runoff may be worth it if one can develop a more real model of catchment behaviour. Soft data abound in our experimental watershed around the world, and this could be a major new direction and use of this information. These soft data measures may move beyond our typical watershed descriptors, into vegetation water use (e.g. Bond *et al.*, 2002) or mean residence time of different groundwater systems contributing to flow in mesoscale watersheds (e.g. Uhlenbrook and Leibundgut, 2002).

Virtual experiments

With the advent of mathematical modelling and computer visualization tools on our desktops, the ability of virtual experiments, as advocated recently by Weiler and McDonnell (2003), may be a way to better improve the dialogue between experimentalist and modeller. The virtual environment provides a hypothesis testing mechanism to elucidate triggers, thresholds and hysteretic relationships in catchment runoff processes. Weiler and McDonnell (2003) developed and implemented a series of virtual experiments whereby the interaction between water flow pathways, source and mixing at the hillslope scale was examined by modeller and experimentalist within a virtual experiment framework. They defined these virtual experiments as 'numerical experiments with a model driven by collective field intelligence'. Virtual experiments are fundamentally different to traditional numerical experiments, since the intent is to explore first-order controls in hillslope and watershed hydrology where the experimentalist and modeller work together cooperatively to develop and analyse the results. Water flux and tracer data are examined jointly to constrain the new conceptualization of how a given hillslope or watershed 'works'. Although not an alternative to field experimentation *per se*, Weiler and McDonnell (2003) argue that virtual experiments may free the experimentalist from the often bewildering array of complexities on his or her hillslope and allow them to use visualization (in addition to the traditional scalar output) as a key interpretive part of the approach and generate working hypotheses for future field experiments. This work was motivated by frustrations that many of us have had in experiments at various hillslopes and watersheds where first-order effects often seem difficult to separate from second- and third-order effects.

Summary

Just as a failure to observe infiltration excess overland flow in forested watersheds led hydrologists to pose an alternative runoff generation concept 50 years ago, non-linearities in internal watershed response to precipitation inputs (thresholds, hysteresis, etc.) compel us to consider new ways to represent the first-order controls of the age, origin and pathway of storm runoff. Certainly, I do not have all the answers for how we craft a post-VSA approach to process description and model structural development. Nevertheless, non-linearity may be a useful framework to sort through the morass of process complexity revealed in the past decades into a straightforward and coherent post-VSA theory. In this way, seemingly different processes, like transmissivity feedback in till soils and pipeflow at a soil–bedrock interface, for example, all become a non-linear threshold process where the switching between states can now be the feature that we might look for in the field, simplifying and focusing our energies with resolve and purpose.

This new direction may alleviate the (futile) search for scale-invariant processes. As Phillips (2003: 9) notes in other disciplines, 'any attempt to identify mechanisms (runoff mechanisms in our case) is doomed to scale-contingent semantic debate over what constitutes a basic mechanism, processes or causal agent and what constitutes a response'. Rather, as one increases in scale, one activates new stores and reservoirs, each with characteristic input–output response, internal mixing dynamics and residence time. The tracer literature is replete with techniques to define these stores. It is now up to the catchment community to rise to the challenge of quantifying these zones (objectively, using physical, chemical and isotopic approaches together) and performing good science to reject ones that are not appropriate. Our work in the coming years should be to go back to the myriad of catchments where we have worked and where we are now working, and perform intercomparison and classification of these non-linear first-order controls on water quantity and quality to inform new generations of watershed models. The task then becomes one of quantifying the first-order controls on connections and disconnections, filling and spilling, etc. across a range of scales and environments. These are the process details that our models will demand if we truly seek explicit water and tracer mass balance in our models. This might be thought of as rationalizing the investment of the first IHD (as we stand on the brink on a new one in the form of Prediction in Ungauged Basins PUB) and returning to some of our geomorphological roots as we seek new ways to match form with process and function.

Finally, although quite critical of VSA theory, I do not want to leave the reader with the impression that near-stream saturated zones are unimportant. Indeed, these areas do produce rapid runoff response to the stream and partitioning of event and pre-event water. In fact, the near-stream saturated zone may be thought of as a prime non-linear reservoir, where behaviour abruptly changes once surface saturation is generated by a rising water table. The point is that the internal description needs rethinking based on what we now know about where water goes when it rains. Catchment hydrologists will need to develop hypotheses from non-linear theory that are testable on the basis of observations in nature. This will not come about via model intercomparison studies or DEM analysis. It will require the dialogue between experimentalist and modeller that has been so lacking in watershed hydrology until now and the practice of hydrology as a true science where field relations are the 'final court of appeal'.

Acknowledgements

These ideas have gelled since co-convening (with Larry Band) the AGU Chapman Conference on State of the Art of Hillslope Hydrology. I would like to express my gratitude to Rick Hooper for many good discussions on this topic over the years and his contribution to my thinking (and my hydrological vocabulary!). I also thank my NSF-project colleagues on EAR-0196381 (Keith Beven, Carol Kendall, Rick Hooper, Jake Peters) who have greatly influenced my thinking on these topics. Many of these ideas fermented during and after productive visits to my lab by Stefan Uhlenbrook and Taro Uchida. My current and recent post docs and PhD students Markus Weiler, Jan Seibert, Jim Freer, Tomas Vitvar, Kellie Vache, Brian McGlynn, Nick Hjerdt, Doug Burns, Mike McHale, Kevin McGuire, Derek Godwin, Willem van Veersveld and Ilja Tromp van Meerveld, are especially thanked for helping most to shape my views on runoff generation. Finally, thanks to Kevin McGuire, Gordon Grant and Tomas Vitvar for their useful comments on an earlier draft of this text.

REFERENCES

Beven KJ. 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences* 5(1): 1–12.

Bond B, Jones J, Moore G, Philips N, Post D, McDonnell JJ. 2002. The zone of vegetative influence on baseflow revealed by diel patterns of streamflow and vegetation water use in a headwater catchment. *Hydrological Processes* 16: 1671–1677.

Burns D, Hooper R, Kendall C, Freer J, McDonnell J, Beven K. 1999. Effect of hillslope flowpaths on subsurface base cation concentration. *Water Resources Research* 34(12): 3535–3544.

Burns DA, McDonnell JJ, Hooper RP, Peters NE, Freer JE, Kendall C, Beven K. 2001. Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). *Hydrological Processes* 15: 1903–1924.

Dunne T. 1978. *Field studies of hillslope flow processes*. In *Hillslope Hydrology*, Kirkby MJ (ed.). Wiley: Chichester; 227–293.

Dunne T. 1983. Relation of field studies and modeling in the prediction of storm runoff. *Journal of Hydrology* 65: 25–48.

Freer J, McDonnell JJ, Beven K, Burns D, Hooper R, Aulenbach B, Kendall C, Peters N. 2002. Understanding the spatial and temporal dynamic contributions of subsurface storm runoff at the hillslope scale. *Water Resources Research* 38(12): 5-1–5-16.

Hewlett JD, Hibbert AR. 1967. Factors affecting the response of small watersheds to precipitation in humid areas. In *Proceedings of 1st International Symposium on Forest Hydrology*; 275–253.

Kirchner J. 2003. A double paradox in catchment hydrology and geochemistry. *Hydrological Processes* 17: 871–874.

Kirchner JW, Hooper RP, Kendall C, Neal C, Leavesley G. 1996. Testing and validating environmental models. *Science of the Total Environment* 183(1–2): 33–47.

McGlynn B, McDonnell JJ. 2003. The role of discrete landscape units in controlling catchment dissolved organic carbon dynamics. *Water Resources Research* 39(4): 3-1–3-18.

McHale M, McDonnell JJ, Mitchell M, Cirmo C. 2002. A field-based study of soil water and groundwater nitrate release in an Adirondack forested watershed. *Water Resources Research* 38(4): 2-1–2-16.

Montgomery DR, Dietrich WE, Torres R, Anderson SP, Loague K. 1997. Piezometric response of a small catchment. *Water Resources Research* 33(1): 91–110.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, I, a discussion of principles. *Journal of Hydrology* 10: 282–290.

Onda Y, Komatsu Y, Tsujimura M, Fujuhara I. 2001. Role of subsurface runoff through bedrock on storm flow generation. *Hydrological Processes* 15: 1693–1706.

Phillips JD. 2003. Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography* 27: 1–23.

Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multi-criteria model calibration. *Water Resources Research* 38(11): 23-11–23-14.

Seibert J, Bishop K, Rodhe A, McDonnell J. 2003. Groundwater dynamics along a hillslope: a test of the steady-state hypothesis. *Water Resources Research* 39(1): 2-1–2-9.

Sivapalan S. 2003. Process complexity at hillslope scale, process simplicity at the watershed scale: is there a consensus? *Hydrological Processes* 17: 1037–1042.

Tromp van Meerveld I, McDonnell JJ. Submitted. Measured non-linearities in subsurface flow: a 147 storm analysis of the Panola hillslope trench. *Water Resources Research*.

Uchida T, Kosugi K Mizumaya T. 2001. Effects of pipeflow on hydrological processes and its relation to landslide: a review pipeflow studies in forested headwater catchments. *Hydrological Processes* 15: 2151–2174.

Uhlenbrook S, Leibundgut C. 2002. Process-oriented catchment modelling and multiple-response validation. *Hydrological Processes* 16: 423–440.

Vitvar T, Burns D, Lawrence G, McDonnell JJ, Wolock D. 2002. Estimation of baseflow residence times in watersheds using the runoff recession hydrograph. *Hydrological Processes* 16: 1871–1877.

Weiler M, McDonnell JJ. 2003. Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology* in press.