HYDROLOGICAL PROCESSES
Hydrol. Process. **15**, 2039–2050 (2001)
DOI: 10.1002/hyp.255

Applying the scientific method to small catchment studies: a review of the Panola Mountain experience

Richard P. Hooper*
US Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA

Abstract:

A hallmark of the scientific method is its iterative application to a problem to increase and refine the understanding of the underlying processes controlling it. A successful iterative application of the scientific method to catchment science (including the fields of hillslope hydrology and biogeochemistry) has been hindered by two factors. First, the scale at which controlled experiments can be performed is much smaller than the scale of the phenomenon of interest. Second, computer simulation models generally have not been used as hypothesis-testing tools as rigorously as they might have been. Model evaluation often has gone only so far as evaluation of goodness of fit, rather than a full structural analysis, which is more useful when treating the model as a hypothesis. An iterative application of a simple mixing model to the Panola Mountain Research Watershed is reviewed to illustrate the increase in understanding gained by this approach and to discern general principles that may be applicable to other studies. The lessons learned include the need for an explicitly stated conceptual model of the catchment, the definition of objective measures of its applicability, and a clear linkage between the scale of observations and the scale of predictions. Published in 2001 by John Wiley & Sons, Ltd.

KEY WORDS hydrochemical models; mixing models; streamflow generation; hillslope hydrology

INTRODUCTION

Although biogeochemical studies of small catchment date back to the 1960s, and physical hydrology studies precede that, we enter the 21st century with only a qualitative understanding of processes operating at the catchment scale, and no consensus on quantitative hydrologic, chemical, or ecological models of catchments. We have learned that disturbances, such as timber harvesting, can alter nutrient cycling (Likens *et al*., 1970), and we have identified potential means to lessen the impact, such as strip cutting and buffer strips. Concepts such as input–output budgets and pool-and-flux diagrams of ecosystems (Likens *et al*., 1977) have been important advances in landscape-scale understanding of ecological processes. The lack of quantitative process understanding, however, means that we are unable to answer reliably common societal questions: How much will the stream chemistry respond to a given reduction in acid deposition? How much will climate change alter the forest ecology? What will be the integrated ecological response to decreased rainfall?

Progress in developing unifying theories in catchment science on which quantitative models could be based has been hindered primarily by the scale of the phenomenon of interest. Our primary observations inside catchments come from instruments, such as lysimeters or permeameters, that measure at the scale of decimetres when we need to understand the landscape at the scale of hectares to square kilometres. Controlled manipulations are exceedingly difficult to perform in the field at relevant scales (for examples of ecosystem-scale manipulations, see Likens *et al*. (1970), Wright *et al*. (1995), and Newton *et al*. (1996)), and even when done suffer from serious artifacts. Laboratory manipulations are still more fraught with artifacts that limit their utility. At its heart, therefore, catchment science remains an observational science, like most of the earth

* Correspondence to: R. P. Hooper, US Geological Survey, 10 Bearfoot Road, Northborough, MA 01532, USA. E-mail: rphooper@usgs.gov

This article is a US Government work and is in the public domain in the United States

Received 1 February 2000
Accepted 7 August 2000

sciences. Success is defined by developing a theory that explains more of the observations in a simpler manner than was done previously.

Computer simulation models can serve as hypothesis-testing tools when traditional experiments cannot be performed. Indeed, the catchment is a particularly well-suited landscape unit for modelling because of its boundary conditions. Outputs are relatively easy to measure, and one of the least ambiguous environmental measurements possible at the landscape scale. Measurement of inputs are complicated by spatial variability of precipitation fields and dry deposition, but are, nonetheless, more straightforward than many other environmental measures, such as hydraulic conductivity and soil chemical properties. Inputs and outputs respectively form the forcing functions and response variables for models. The model is composed of hypothesized mechanisms, and the model predictions are compared with the output signal. In catchment studies, simulation models are proposed and accepted when the predictions satisfactorily match the observed hydrograph and chemographs.

This approach may be suitable when the goal of the exercise is simply a prediction of system response to a stress. However, these efforts do not serve as hypothesis tests for a number of reasons. First, a hypothesis can only be rejected definitively. It is a logical fallacy to assume that a hypothesis can be accepted because there may be alternate explanations for the observed behaviour. Agreement between observations and predictions is only a necessary, not a sufficient, condition for the hypothesis to be correct. Ideally, the set of all models that are consistent with the data should be identified, rather than simply noting the agreement between one model and the data, as is commonly done. More importantly, we have inverted the standard design of a scientific experiment. In a standard experimental design, the null hypothesis is constructed to state our belief in the absence of data to the contrary. Data are collected to reject the null hypothesis, and we believe that we have learned something when we can reject this hypothesis: the experiment was a success. However, in our use of simulation models, success is defined by accepting the model, i.e. accepting the null hypothesis that the mechanisms contained in the model are 'correct'. Because the test is structured in this way, we must be concerned with the ability of the test to reject an incorrect model. Statistically, this is known as the power of the test.

Catchment simulation models frequently contain parameters that cannot be measured directly, either because of spatial aggregation within the model or because they are conceptual rather than physical parameters. The presence of parameters that must be fit to the data reduces the power of the hypothesis test to reject incorrect hypotheses. More complex models (generally considered to be more 'realistic') usually have more unmeasurable parameters, further lowering the power of the test. Kirchner *et al*. (1996) provide a compelling example of the importance of power in computer simulation models used to test hypotheses.

Given the nature of catchment studies and this modelling conundrum, how can progress be made? In this paper, I review the results of a formal structural analysis of the Birkenes model, a simple catchment hydrochemical model. Upon finding that the simulation model indeed did have low power as a hypothesis test, my colleagues and I developed a simple mixing model that serves as a quantitative conceptual model. This model was applied iteratively and refined in a series of studies conducted at the Panola Mountain Research Watershed (PMRW) and has helped to focus field efforts to discern streamflow generation mechanisms. I review these studies to explore the relation between field observations and simulation modelling and to elucidate general principles that can be applied to make catchment studies more effective.

SIMULATION MODELS AND INFORMATION CONTENT OF DATA

One way to examine the power of simulation models is to 'invert' the model. That is, instead of specifying a set of parameters and running the model to obtain predictions, the model is run many times in order to determine the set of all parameters that are consistent with the data. Ideally, there is a single parameter set that explains the observations 'significantly' better than all others do. This implies that there is sufficient information in the calibration data set to determine a unique parameter set. Such a model has high power

because parameter values cannot be varied to compensate for errors in the model. In general, watershed models do not exhibit this behaviour. The set of parameters that yield virtually the same model performance can be quite large (Beven and Binley, 1992; Hornberger and Boyer, 1995).

Model inversion is not done routinely for a number of reasons. Even the simplest models cannot be inverted analytically, so numerical optimization routines must be employed. A quantitative objective function (such as the sum of squared errors) must be specified to define model performance; all such functions measure only some aspect of the model fit and may not be wholly satisfactory. The dimensionality of the optimization procedure increases with each parameter to be determined. Until recently, computing power was a significant limitation to our ability to invert models because of the high dimensionality of the parameter space that must be searched by the optimization routines. Finally, for these numerical optimization routines to perform correctly, the error structure of the model residuals (such as autocorrelation among observations and unequal variance in the residuals with respect to an independent variable) must be considered (e.g. Sorooshian and Dracup, 1980). Thus, performing this analysis is burdensome.

Despite these limitations, the hydrologic module of the Birkenes acidification model (Christophersen and Wright, 1981) was subjected to a inverse analysis by de Grosbois *et al*. (1988) and Hooper *et al*. (1988) using data collected from the Birkenes catchment in Norway. The hydrologic model is a six-parameter, two linear-reservoir model that operates on a daily time step. By using rainfall and a conservative tracer, the dimensional and rate parameters could be determined, but the routing parameter determining whether the water exiting the upper reservoir went directly to the stream or to the lower reservoir could not be determined from these data. This parameter is critical in determining the chemical response of the stream to acid deposition. Potentially, incorporating a reactive solute, such as sulfate, into the calibration data set could make the routing parameter observable. However, a mathematical model of a reactive solute requires additional parameters, such as exchange coefficients with the soil and equilibrium constants, that must be fit. The critical question is whether the information content of the reactive chemical data is greater than the information required to determine the new parameters. A preliminary analysis of this question indicated that there was *not* sufficient information gained to improve the identifiability of the hydrologic routing parameter (Stone and Seip, 1990).

Many authors have found similar results. Beven and Binley (1992), using the Generalized Likelihood Uncertained Estimation (GLUE) procedure, showed that many parameter sets were equally likely to provide an acceptable simulation of the hydrograph using TOPMODEL. This work has been extended to determine the value of longer time series (Freer *et al*., 1996) and of water-table elevations (Lamb *et al*., 1998) in constraining parameters. Although some improvement was found, many parameters remain poorly defined. Ultimately, inferring within-catchment processes from the outflow hydrograph and chemographs alone cannot be done uniquely because many alternate mechanisms can produce the observed output (Buttle, 1994). Procedures, such as GLUE, provided the analytical tools to demonstrate quantitatively how non-unique the solutions are.

To use the Birkenes model as a hypothesis testing tool, substantial changes were needed: either increase the information that could be brought to bear on determining the parameters or decrease the information required by reducing the number of parameters. The critical conclusion was that there was no point in continuing to use this model to test acidification mechanisms because it had insufficient power to reject an incorrect hypothesis. Bear in mind that this was true despite the fact that the model could fit the data.

These structural analyses have been criticized as being overly formal and not appropriate when applied to 'trans-scientific' questions (Weinberg, 1978), such as prediction of the catchment response to different loads of acid deposition. These criticisms miss the point of the exercise. This is not a narrow hypothesis test with a simple 'accept' or 'reject' answer. Rather, it indicates the limitation of the model as a hypothesis-testing tool and suggests that if the point of the exercise is to evaluate the validity of the mechanisms represented in the model, this is not the appropriate tool to use. Furthermore, although the question of the future response of a catchment to a stress may be 'trans-scientific' (i.e. not directly amenable to scientific investigations), the point of modelling in the first place was to provide a more reliable prediction than could otherwise be done because the models were based upon the 'correct' mechanisms. (Often, an explicit benchmark against which

to compare model performance is never stated. See Kirchner *et al*. (1996) for further discussion.) Assessing whether the mechanisms are correct (or at least consistent with the data) is certainly a 'scientific' exercise.

DEFINING A CONCEPTUAL MODEL

Because we cannot test mechanisms by using a traditional simulation model, it is necessary to state a testable hypothesis in another manner. An inductive 'leap' is required, because there is no obvious next step after discarding the model. Recognizing that we wanted to define hydrologic flowpaths through a catchment in an objective manner, based upon data, we asked the question: could streamwater be composed by mixing varying proportions of soil and groundwater observed within the catchment? This is the simple, commonly used mixing model of source 'end members' to compose a solution that has been employed in a variety of situations, including estuarine mixing problems, hydrograph separation, and many others. To simplify the mathematics of this question (and to avoid additional model parameters), we restricted the question to considering only conservative solutes.

This conceptual model has a number of desirable properties. First, there is an explicit linkage between internal catchment measurements (soil and groundwater solution chemistry) and a catchment-scale observation (stream chemistry). Second, a measurement scale—the variation in solute concentration at the catchment outlet—arises naturally from the model statement. This provides a metric both for measuring the performance of the model (how well does the model explain this variation) and against which to compare the variability of end-member solutions. When we first proposed this model, many soil scientists were skeptical of its viability because of the high variability of soil solutions. Soil solutions, indeed, are variable in time and space; however, if this variation is small relative to variation in stream chemistry, and if the solutions are distinct chemically from one another, they may serve as effective end members. Finally, by focusing on variation in stream chemistry, the limits of this conceptual model are clear. Only those catchment processes that influence the stream chemistry at the outlet can be included; those that are not expressed in this way are not observable. This is an explicit recognition that we are building the model on the information content of the stream chemistry signal.

Site description

The PMRW, a 41 ha basin, is located in the Piedmont physiographic province, approximately 25 km southeast of Atlanta, Georgia, USA (Figure 1). The basin is underlain primarily by granodiorite emplaced about 320 Ma in a host rock of biotite–plagioclase gneiss that contains amphibolite and K-feldspar (Higgins *et al*., 1988). There is a 3 ha granodiorite outcrop in the headland of the catchment. Soils at PMRW are classified as entisols near the base of the outcrop and as ultisols on hillslopes and ridgetops. Soils have developed to a thickness of 1 to 2 m, and saprolite underlies the soils to a depth of 0 to 5 m below land surface. The basin, with the exception of the outcrop, is completely forested, with mixed southern hardwoods covering 70% of the forested area, and pine covering the remainder. The climate in the area is classified as warm temperate subtropical, with an annual average temperature of 16 °C (Hodler and Schretter, 1986). The long-term annual average precipitation in the area is 1·24 m; annual average runoff is 30% of precipitation (Carter and Stiles, 1983), but this percentage varies widely from year to year as a function of both rainfall amount and timing. The stream at the catchment outlet (the lower gauge) and at the southeastern tributary streamflow gauge is perennial; flow at the upper gauge on the southwest tributary is ephemeral, with sustained flows between storms occurring only in the winter.

Application of the conceptual model to PMRW

At PMRW, six solutes—calcium, magnesium, sodium, sulfate, silica, and alkalinity—were thought to be conservative and were used as tracers (Hooper *et al*., 1990). Fifteen mixing diagrams, such as shown in

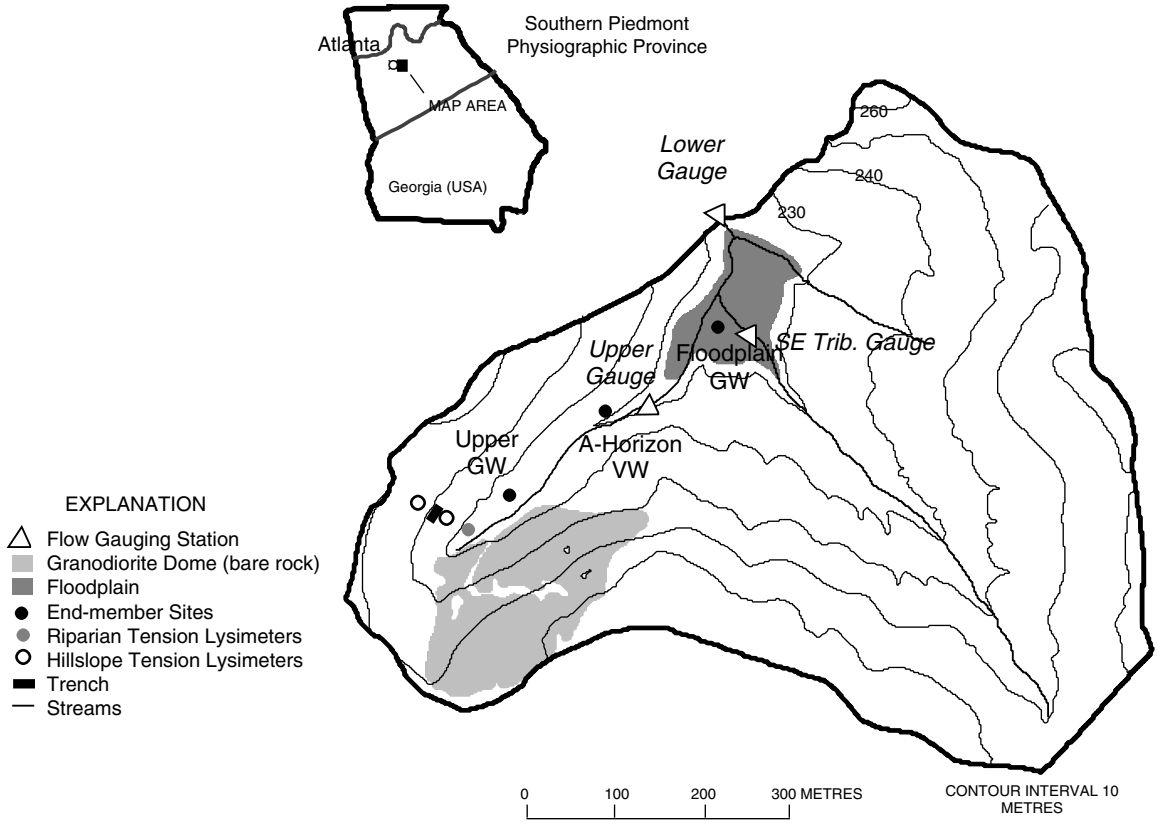

Figure 1. Map of PMRW showing location of gauges, lysimeters, wells, and hillslope trench

Figure 2, can be constructed for all pairs of conservative solutes. Groundwater and soilwater solutions are plotted on the same diagram as the stream solutions. Viable end-member solutions have concentrations more extreme than the stream, have a low variability relative to the stream variation, and are distinctly different from other end members. In Figure 2, the end-member solutions are represented by their median; variability in concentration is indicated by the quartiles, drawn as bars in the figures. The three end members identified at PMRW were groundwater from the floodplain ('floodplain GW' in Figure 1), groundwater from the upper part of the catchment ('upper GW'), and vadose zone water collected at the bottom of the A-horizon of the soil ('A-horizon VW'). As can be seen in Figure 2, the variation in the end-member solutions is much less than the variation in the stream chemistry and the end members are distinct from one another.

If the median concentration is used as the representative concentration for the mixing model, any stream sample that lies in the interior of the triangle formed by connecting the three end-member medians can be explained as a mixture of the three end members. Although this triangle did not contain all the data points, the mixing model reproduced the measured chemistry very well. The coefficient of determination r^2 ranged from 0·82 for calcium to 0·97 for alkalinity. The hydrograph separation implied by the mixing model was also plausible. Floodplain GW dominated baseflow, particularly during the drier summer months; upper GW was the dominant source during the wetter winter months; and the A-horizon VW contributed only during storms.

Although this mixing model apparently fits the data, our confidence in it stems from matching the outflow chemograph. As noted above, the chemograph does not uniquely determine the controlling processes. However, unlike the Birkenes model, which contained two 'conceptual' reservoirs, the end members define actual

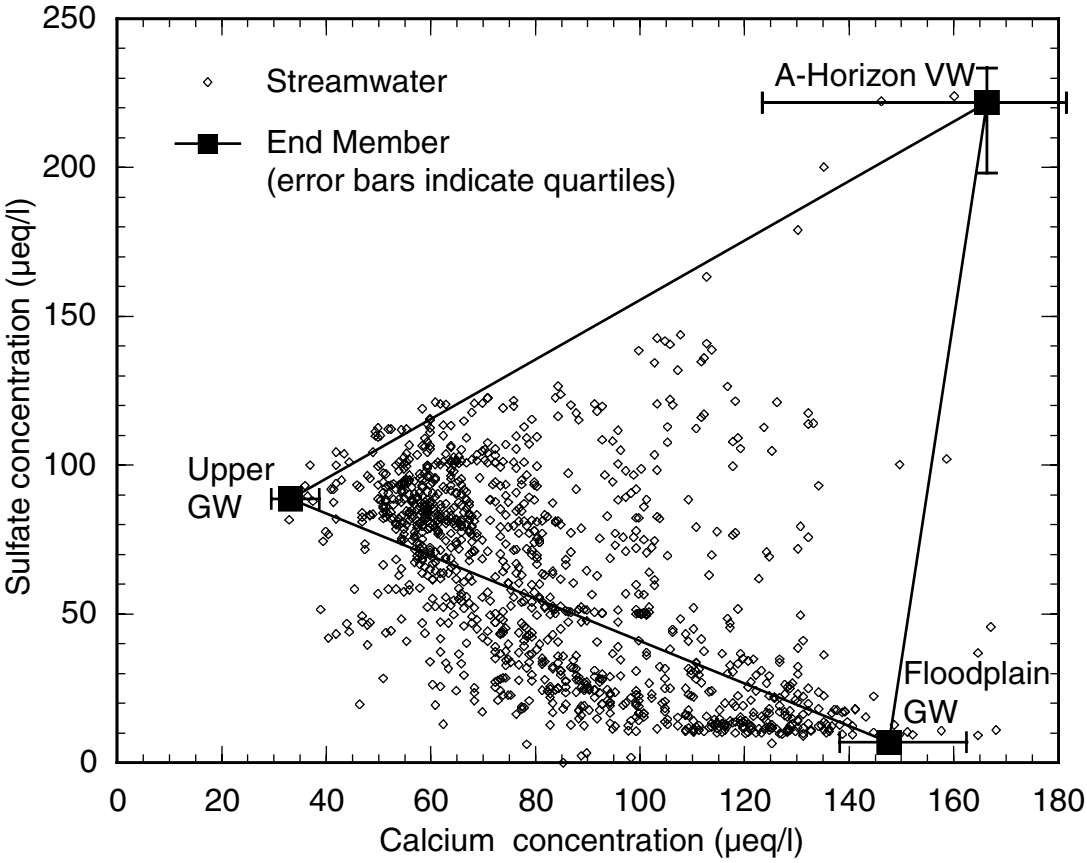

Figure 2. Mixing diagram constructed using data collected between 1 October 1985 and 30 September 1988

physical environments. Hooper *et al*. (1990) laid out a number of observations necessary to test the validity of the mixing model. These included sampling to determine the spatial extent of the end members and sampling along hydrologic flowpaths to determine whether discrete end members existed or if there is a continuum of soil solution chemistries. Furthermore, hydrometric observations could be used to test the flowpaths implied by the hydrograph separation. The critical point is that the proposed conceptual model is testable with internal catchment observations.

REFINING THE CONCEPTUAL MODEL

End-member mixing analysis (EMMA), as proposed by Christophersen *et al*. (1990) and Hooper *et al*. (1990), solved an over-determined system of equations, i.e. more tracers were used than were necessary to define the end members. For three end members only two tracers are required to define mixing proportions, whereas at PMRW six solutes were used. By exploring the geometry of the mixing problem, Christophersen and Hooper (1992) found a conceptual error in the original solution technique and proposed modifications that were more consistent with the intended hypothesis test. Originally, stream samples were projected into the sub-space defined by the end members to assess whether they could be explained as a mixture. Because the stream chemistry is known, and the end members are being tested, this is backwards. Christophersen and Hooper (1992) developed an approach to project the end members into the sub-space defined by the

stream samples for evaluation of the mixing model. The results of this analysis indicated that the projections of the A-horizon VW end member had a significantly higher sodium concentration and a significantly lower alkalinity concentration than the measured end members. These differences raised a question as to whether the end member was truly the A-horizon soil solution, runoff from the granite outcrop, or throughfall. Evidence supporting the contention that the soil solution was the end member included the high degree of regulation of streamwater chemistry observed just metres below the granite outcrop by Shanley and Peters (1993).

Hooper and Christophersen (1992) used this modified technique to identify parameters for a long-term acidification model, MAGIC or the Model of Acidification of Groundwater in Catchments (Cosby *et al.*, 1985a,b). Both the hydrologic routing parameters and soil chemical and dimensional parameters of the two-reservoir version of MAGIC (Jenkins and Cosby, 1989) were traditionally determined by matching the stream chemistry. Hooper and Christophersen (1992), instead, fit the soil parameters directly to the soil environment from which the end member arose, and determined the routing parameters by determining the proportion of each end member required to compose the annual average stream chemistry, consistent with the annual time step used by MAGIC. This parameterization approach was far more constrained than the traditional one, because the interactions among the parameters were limited by dividing the fitting process into the soil parameters and the hydrologic parameter.

The results of this analysis indicated that two factors were most important in determining the stream chemical response to acid deposition: the shallow A-horizon that had a limited capacity to retain sulfate and the low alkalinity upper-GW end member that rapidly acidified once sulfate was transported from the A-horizon. Neither of these factors would have been apparent without the conceptual mixing model. Furthermore, because actual soil environments were used in the model, the model predictions are testable. For example, the surficial soil horizon was predicted to be rapidly approaching sulfate saturation, and that sulfate would begin migrating down the soil and acidify the upper GW soil solution. A plot study could be designed to accelerate sulfate saturation by adding sulfate to test the predicted dynamics.

This application of the conceptual model, however, was based on the assumption that the end members came from an environment that had a large spatial extent. For example, the upper GW end member was assumed to exist everywhere in the catchment outside the floodplain. Therefore, choosing the one groundwater solution that was the best end member was not appropriate: it was necessary to consider all observations within the end-member environment to determine a 'typical' concentration. Another assumption of EMMA is that the concentrations of end members are temporally invariant. Although there were limited observations of short-term variability in the A-horizon end member, no intensive sampling was performed of groundwater solutions. We assumed that variation in stream chemistry during storms could be explained by mixing of temporally invariant groundwater solutions; however, it is possible that groundwater solutions themselves could vary in concentration during events. A second iteration of field work was planned in which the spatial extent of the end members was tested and their temporal variability assessed.

TESTING THE CONCEPTUAL MODEL

The original EMMA performed at PMRW used data collected between 1 October 1985 and 30 September 1988, or Water Years 1986 through 1988. (The USGS defines a water year as beginning on 1 October and ending on 30 September. Water Year 1986 began 1 October 1985 and ended 30 September 1986.) Continued monitoring of the A-horizon end member indicated that sulfate concentrations declined markedly between 1988 and 1990 (Figure 3). By 1991, the sulfate concentration had reached about one-half the pre-1989 concentration, and the downward trend in concentration continued at a slower rate. Calcium concentration also declined to approximately one-half of the pre-1989 concentration. Other solutes diluted slightly, but sulfate and calcium were the two most strongly affected.

This change in concentration provided the strongest test to date of the conceptual model: Would the stream reflect this large change in concentration? Because the A-horizon end member contributes to streamflow

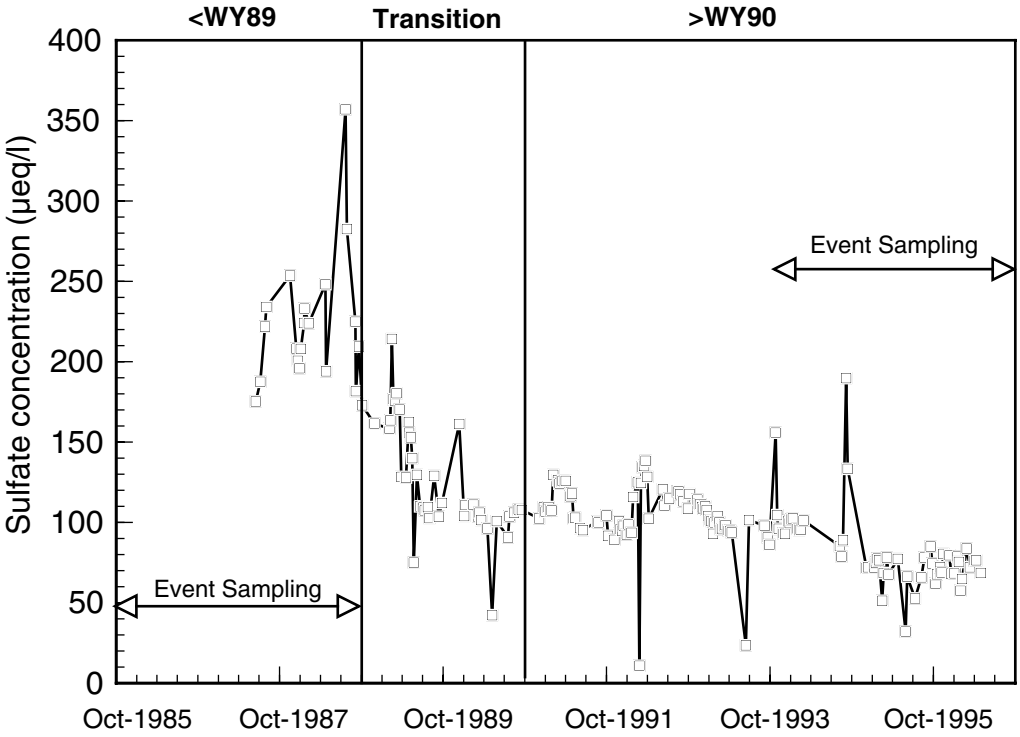

Figure 3. Time series of sulfate concentrations from A-horizon vadose water end member

significantly only during storms, storm sampling was re-instituted in October 1993 after having been suspended in 1989. As can be seen in Figure 4, the stream chemistry did change in a manner consistent with the mixing model hypothesis. Peak sulfate and calcium concentrations in the stream were much lower after 1991 than they were before 1989. This provides the strongest evidence to date that the A-horizon is controlling the stream chemistry.

Testing the spatial extent of the end members proved more difficult. On the hillslopes above the riparian zone at PMWR, zero-tension water could not be found reliably. Most piezometers were dry, even when located in surface hollows, and zero-tension lysimeters collected samples only at shallow depths in large events. At other times, they were dry. Therefore, three nests of tension lysimeters were installed in 1991 and sampled approximately monthly for 2 years. Two nests were located on the hillslope opposite the granite outcrop and a third in the riparian zone (Figure 1). If the conceptual model is accurate, the shallow lysimeters should have a chemistry similar to the A-horizon end member and the deep lysimeters should have a chemistry similar to the upper GW end member.

Samples of the saturated zone during storm events was made possible by a 20 m long trench excavated to bedrock. This trench was installed in 1995 as part of a larger study of hillslope hydrology (McDonnell *et al*., 1996) to isolate a hillslope segment. The trench was located just above the lower of the two hillslope lysimeter nests, approximately 50 m upslope of the stream (Figure 1). The trenchface was divided into 2 m segments, and water samples were collected during storms in the winter and spring of 1996. These samples are the first observation of mobile water at PMRW from the saturated zone during storms. The chemistry of these samples was hypothesized to be similar to the upper groundwater end member.

The chemistry of these solutions was very different from what was hypothesized (Figure 5). Samples from the shallow tension lysimeters had a much higher concentration than the post-WY90 A-horizon end member. This difference may be because these tension waters were qualitatively different than the zero-tension water

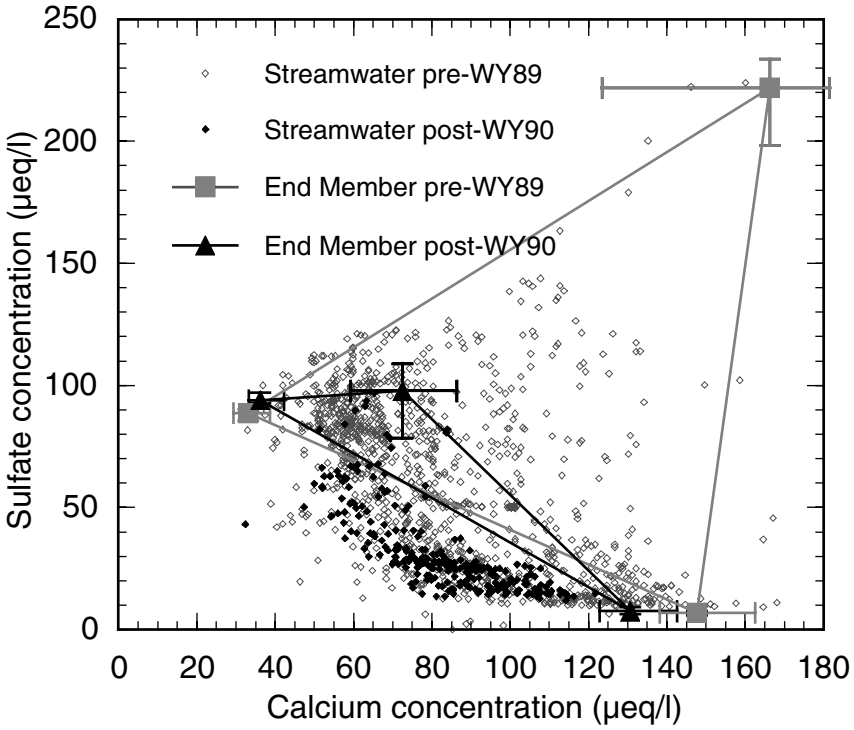

Figure 4. Mixing diagram indicating differences in end members and streamwater chemistry during two periods, prior to Water Year 1989 and after Water Year 1990

sampled for the end member. However, the three deepest hillslope lysimeters, were still much higher in concentration than either the A-horizon or the upper GW end member. There also was a large difference between the hillslope tension lysimeters and the riparian zone ones. The deeper riparian tension lysimeters had a chemistry more similar to the upper GW end member. The 120 cm deep lysimeter, in fact, would be a better end member than the upper groundwater solution.

The zero-tension trenchface samples were similar in concentration to the deeper hillslope tension lysimeter samples. This finding suggests that the samples from the tension lysimeters were representative of the chemistry in the hillslope saturated zone. Nonetheless, the trenchface samples were chemically distinct from both the A-horizon end member and the upper-GW end member. The dominant cation and anion for the trenchface samples were calcium and sulfate respectively; the streamwater is a sodium bicarbonate water. This chemical difference exists despite the proximity of the trench to the stream.

Also shown in Figure 5 are samples from the southeastern tributary. Overall, their chemistry is similar to those from the lower gauge. In particular, the same low concentrations of calcium and sulfate are evident. Thus, the dilution observed during events is not solely a function of the granite outcrop along the southwestern tributary. There is an apparent contradiction. The conceptual model was supported when the change in one controlling end member was reflected in the stream, yet hillslope solutions had chemistry very different from that which was hypothesized. The resolution to this problem is that the stream chemistry is reflecting relative contributions of different parts of the riparian zone (the floodplain GW and upper GW end members, primarily), but does not reflect the hillslope water, at least at the time scale of storms.

The implications of this finding for catchment modelling are profound. We cannot use the stream chemistry to infer a structure for the hydrologic model as Hooper and Christophersen (1992) did because the hillslope

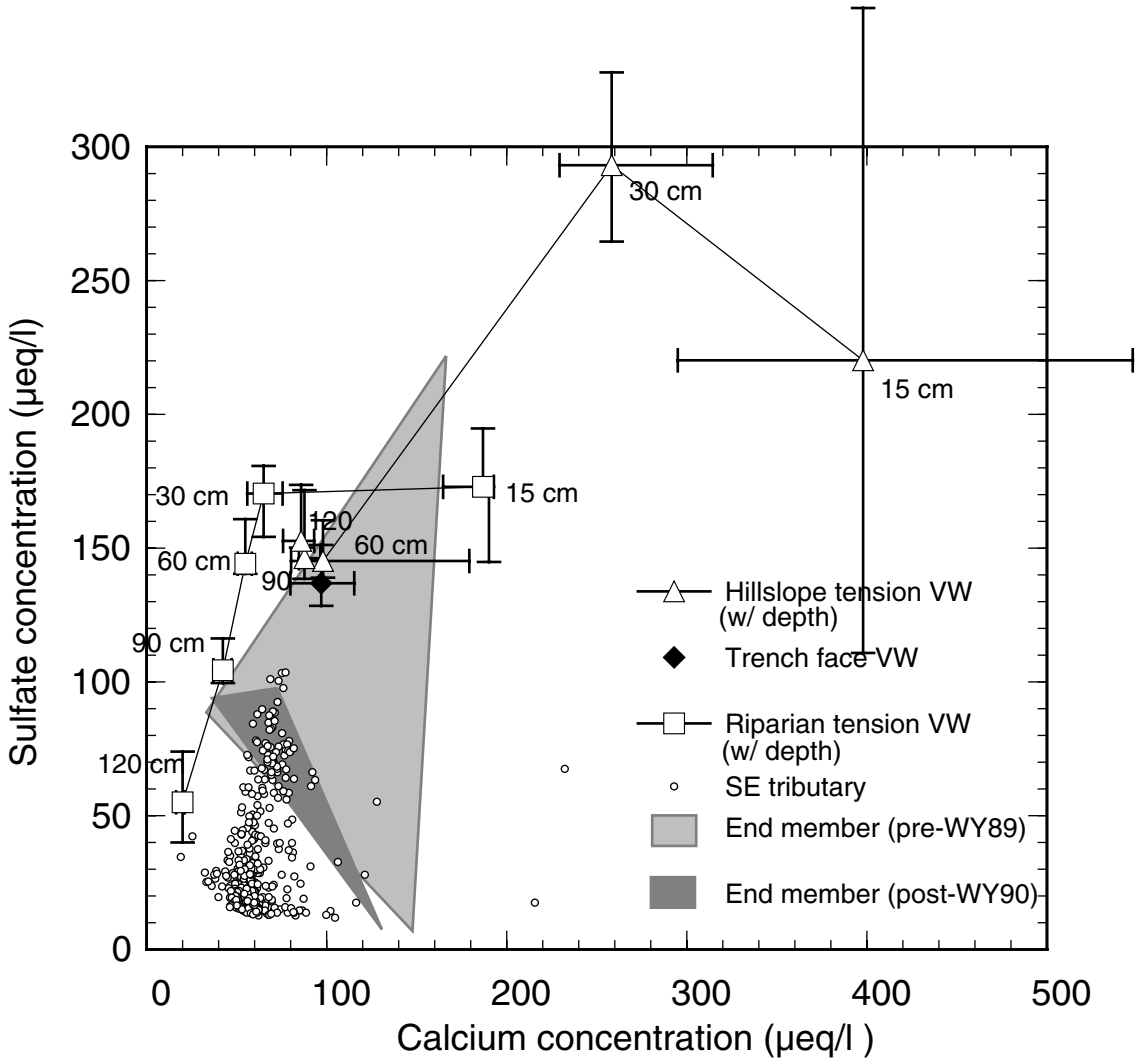

Figure 5. Mixing diagram comparing hillslope, lysimeter, and streamwater samples from southeastern tributary with hypothesized end members

area of the catchment, the largest landform in the catchment, is not chemically expressed in the stream. This finding calls into question the use of stream chemical dynamics to calibrate predictive hydrochemical models, because those dynamics provide no indication of the chemical processes occurring outside the riparian zone. At PMRW, streamflow chemical dynamics largely reflect the relative contribution of different parts of the riparian area, and not the workings of the catchment as a whole.

TOWARDS IMPROVED CATCHMENT UNDERSTANDING

Another iteration of fieldwork and modelling has been performed, a hypothesis disproven, and more questions raised. Has progress been made? I believe that the answer is 'yes'. Much of the variation in stream chemistry, at least at PMRW, reflects differing proportions of water coming from the upper and lower portions of the riparian zone. From this point of view, we have learned what can be learned from frequent stream sampling

during events. The largest part of the catchment, the hillslope area, is not chemically expressed in the stream during events, yet certainly this area supplies water (and solutes) to the stream. Perhaps there are chemical reactions that occur as the hillslope water enters the riparian zone that radically change the solution chemistry. Perhaps the hillslope water is volumetrically insignificant compared with water resident in the riparian zone. Perhaps there is a delay in the translation of water from hillslope to riparian zone, and the majority of transport occurs not during the storm, but between storms when there is also more opportunity for chemical transformations.

These results suggest that field studies shift their focus from the chemical dynamics that occur during storms to explore the connection between the hillslope and riparian portions of the catchment. How are they connected? How is water distributed between events? Does stream volume aggrade gradually with increasing upstream area or are there discrete points of large increases? Can the chemically distinct hillslope water be discerned in the alluvial aquifer? One hypothesis is that the longer the residence time in the alluvial aquifer, the less the hillslope chemistry is expressed in the stream. This hypothesis is being tested in three catchments, from one with a long residence time (PMRW), to a moderate residence time (Sleepers River, Vermont, USA) to a wet catchment with short residence time (Mai Mai, New Zealand).

Through an iterative application of a simple conceptual model to a single catchment, we have been able to focus our field activities towards answering increasingly specific questions of hillslope hydrology and geochemical evolution. Although this paper focuses on the chemical aspects of a mixing model, clearly hydrometric approaches must also be employed to test the flowpaths implied by the mixing model. McDonnell *et al*. (In press) demonstrate a refinement of the perceptual model of streamflow generation at Mai Mai, New Zealand, from repeated field studies. Common to both the Mai Mai and PMRW experiences was the need to link internal observations to catchment-scale behaviour and to develop a measurement scale that enables us to distinguish signal from noise.

Although important limitations exist on how we measure internal catchment variables, catchment scientists have the ability to generate data series that were unimaginable 10 years ago. We still must decide, however, what to measure and where to measure it. Rigorous and iterative application of a quantitative conceptual model will permit us to deploy our tools in an effective way to advance our understanding of catchments.

REFERENCES

Beven K, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.

Buttle JM. 1994. Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Progress in Physical Geography* **18**: 16–41.

Carter RF, Stiles HR. 1983. Average annual rainfall and runoff in Georgia, 1941–70. *Hydrology Atlas 9*. Georgia Geological Survey: Atlanta, GA, USA.

Christophersen N, Wright RF. 1981. Sulfate budget and a model for sulfate concentration in streamwater at Birkenes, a small forested catchment in southernmost Norway. *Water Resources Research* **17**: 377–389.

Christophersen N, Hooper RP. 1992. Multivariate analysis of stream water chemical data: the use of principal components analysis for the end-member mixing problem. *Water Resources Research* **28**: 99–107.

Christophersen N, Neal C, Hooper RP, Vogt RD, Andersen S. 1990. Modelling streamwater chemistry as a mixture of soilwater end members—a step towards second-generation acidification models. *Journal of Hydrology* **116**: 307–320.

Cosby BJ, Hornberger GM, Galloway JN, Wright RF. 1985a. Modeling the effect of acid deposition: assessment of a lumped parameter model of soil water and streamwater chemistry. *Water Resources Research* **21**: 51–63.

Cosby BJ, Wright RF, Hornberger GM, Galloway JN. 1985b. Modeling the effect of acid deposition: estimation of a long term water quality responses in a small forested catchment. *Water Resources Research* **21**: 1591–1603.

De Grosbois E, Hooper RP, Christophersen N. 1988. A multisignal automatic calibration methodology for hydrochemical models: a case study of the Birkenes model. *Water Resources Research* **24**: 1299–1307.

Freer J, Beven K, 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**: 2161–2173.

Higgins MW, Atkins RL, Crawford TJ, Crawford RF, Brooks R, Cook R. 1988. The structure, stratigraphy, tectonostatigraphy, and evolution of southernmost part of the Appalachian region. *USGS Professional Pap. 1475*, Reston, VA, USA.

Hodler TW, Schretter HA. 1986. *The Atlas of Georgia*. University of Georgia Press: Athens, GA, USA.

Hooper RP, Christophersen N. 1992. Predicting episodic stream acidification in the southeastern United States—combining a long-term acidification model and the end-member mixing concept. *Water Resources Research* **28**: 1983–1990.

Hooper RP, Stone A, Christophersen N, de Grosbois E, Seip HM. 1988. Assessing the Birkenes model of stream acidification using a multisignal calibration methodology. *Water Resources Research* **24**: 1308–1316.

Hooper RP, Christophersen N, Peters NE. 1990. Modelling streamwater chemistry as a mixture of soilwater end members—an application to the Panola Mountain Catchment, Georgia, USA. *Journal of Hydrology* **116**: 321–343.

Hornberger GM, Boyer EW. 1995. Recent advances in watershed modeling. *Reviews of Geophysics* **33**(Part 2, suppl. S): 949–957.

Jenkins A, Cosby BJ. 1989. Modeling surface water acidification using one and two soil layers and simple flow routing. In *Regional Acidification Models*, Kamari J, Brakke DF, Jenkins A, Norton SA, Wright RF (eds). Spring-Verlag: New York; 253–266.

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

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

Likens GE, Bormann FH, Johnson NM, Fisher DW, Pierce RS. 1970. The effect of forest cutting and herbicide treatment on nutrient budgets in the Hubbard Brook watershed-ecosystem. *Ecology Monographs* **40**: 23–47.

Likens GE, Bormann FH, Pierce RS, Eaton JS, Johnson NM. 1977. *Biogeochemistry of a Forested Ecosystem*. Springer-Verlag: New York; 146 pp.

McDonnell JJ, Freer J, Hooper RP, Kendall C, Burns DA, Beven KJ, Peters N. 1996. New method developed for studying flow on hillslopes. *EOS, Transactions of the American Geophysical Union* **77**: 465.

McDonnell JJ, McGlynn B, Brammer D. In press. An evolving perceptual model of hillslope flow in a steep forested humid catchment: a review of the Mai Mai catchment. *Journal of Hydrology*.

Newton RM, Burns DA, Blette VL, Driscoll CT. 1996. Effect of whole catchment liming on the episodic acidification of two Adirondack streams. *Biogeochemistry* **32**: 299.

Shanley JB, Peters NE. 1993. Variations in aqueous sulfate concentrations at Panola Mountain, Georgia. *Journal of Hydrology* **146**: 361–382.

Sorooshian S, Dracup JA. 1980. Stochastic parameters estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resources Research* **16**: 430–442.

Stone A, Seip HM. 1990. Are mathematical models useful for understanding water acidification? *Science of the Total Environment* **96**: 159–174.

Weinberg AM. 1978. Trans-science. *Nature* **273**: 93.

Wright RF, Brandrud T-E, Clemensson-Lindell A, Hultberg H, Kjonaas OJ, Moldan F, Persson H, Stuanes AO. 1995. NITREX project: ecosystem response to chronic additions of nitrogen to a spruce-forested catchment at Gårdsjøn, Sweden. *Ecology Bulletin* **44**: 322.