On the use of multiple criteria for *a posteriori* model rejection: Soft data to characterize model performance

Kellie B. Vaché and Jeffrey J. McDonnell
Department of Forest Engineering, Oregon State University, Corvallis, Oregon, USA

John Bolte
Department of Bioengineering, Oregon State University, Corvallis, Oregon, USA

Received 21 September 2004; accepted 6 October 2004; published 4 November 2004.

[1] Land surface hydrologic models are commonly evaluated based upon the degree of correspondence between measured and modeled discharge. In this paper we illustrate significant shortcomings associated with the simple discharge based evaluation strategy. A standard conceptual hydrologic model is applied within a Monte Carlo framework to two catchments representing significantly different hydrologic regimes. Time source hydrograph separations are derived, in addition to modeled discharge, and used to more completely characterize model functioning, across the entire *a priori* parameter distribution. The inclusion of hydrograph separation results in improved characterization of parameter uncertainty, and in one of the cases, complete model rejection. *INDEX TERMS:* 1806 Hydrology: Chemistry of fresh water; 1836 Hydrology: Hydrologic budget (1655); 1832 Hydrology: Groundwater transport. **Citation:** Vaché, K. B., J. J. McDonnell, and J. Bolte (2004), On the use of multiple criteria for *a posteriori* model rejection: Soft data to characterize model performance, *Geophys. Res. Lett.*, *31*, L21504, doi:10.1029/2004GL021577.

1. Introduction

[2] Watershed modeling is an activity fundamental to the hydrological sciences for both hypothesis testing and prediction related to fluxes of water in the environment. Much recent thought has been given to model parameter identification and uncertainty [*Beven*, 2001]. Scatter plots of model efficiency versus parameter values are now a well-established method of depicting parameter identifiability [*Dunn et al.*, 2003]. These *a posteriori* parameter distributions are developed through Monte Carlo analysis, where an ensemble of models is generated to represent different parameter vectors and efficiencies. The efficiency indicates the model's similarity to measured output, often focused upon stream discharge. The scatter plots focus attention on the parameter uncertainty due to equifinality, common to all conceptual rainfall runoff models [*Beven*, 2001]. Equifinality is defined by the fact that many different parameter vectors will often provide equally efficient simulations with respect to a specific single measure of performance. While data-based modeling procedures [*Jakeman and Hornberger*, 1993] seek to reduce the number of parameters to a value supported by the data, they lose one of the major benefits of the conceptual approach – a realistic definition of plausible flowpath mechanics.

0094-8276/04/2004GL021577

[3] New toolkits to examine scatter plots have been recently developed [*Wagener et al.*, 2003]. However, few diagnostic tools utilize directly our process understanding of the rainfall-runoff process. Recent work has suggested that process information can be used in model calibration via "soft data". *Seibert and McDonnell* [2002] define soft data as data from the experimentalist that cannot be used as hard numbers in a traditional model calibration sense. As such, soft data are qualitative information derived through experimentation and post experiment data manipulation. While fuzzy, these soft data often describe the system in ways that map to field scientist's process understanding. *Seibert and McDonnell* [2002] used soft data in *a priori* analyses, where the modeler provided evaluation rules related to model state (e.g., % saturated area) or parameters (soil depth) and the experimentalist provides fuzzy descriptions of these features. This communication can result in the incorporation of process understanding into the modeling exercise. However, parameter identification through inversion is by definition *a posteriori*. While this process clearly benefits from improved *a priori* parameter estimates, there remains significant potential for more complete comparisons between model output and process understanding to significantly improve model realism.

[4] To date, much of the focus on *a posteriori* interpretations has been on how well identified a particular parameter might be, and how this translates into parameter uncertainty [*Freer et al.*, 1996]. In this paper, we extend that idea and focus on model evaluation through further classification of the set of acceptable (or behavioral) models using field-based process understanding of basin dynamics. We argue that acceptance, in the case of a conceptual model where runoff pathways are specifically defined, should be based upon the degree to which simulated runoff pathways correspond to process understanding. This process understanding has been very difficult to capture in a single integrated measure - like a stream hydrograph. One expression of process dynamics hitherto not linked to model evaluation is the isotope-based hydrograph separation. Since the late 1960s a wide variety of papers have reported the time-source components of storm hydrographs [*Burns*, 2002]. For catchments in humid zones, the hydrograph split of event vs pre-event water (the two time source components) averages 25% event water and 75% pre-event water [*Buttle and McDonnell*, 2004]. Alternatively, studies reporting hydrograph separations from semi-arid zones [*Sandstrom*, 1996] suggest a dominance of event water. In this paper, we assert that the hydrograph separation literature

is now mature enough to use these general observations as useful process-based criteria for model evaluation. Like a hydrograph, the time source separation is an integrated measure of basin response. Unlike hard data represented by a flow hydrograph, hydrograph separation data might be considered a form of soft data, the incorporation of which requires a degree of subjectivity. This paper demonstrates the utility of the combination of both discharge-based efficiency and hydrograph separation to more clearly differentiate between behavioral and non-behavioral simulations. It builds upon the work of *Seibert and McDonnell* [2002] through the presentation of an explicit tracer model and documentation of the utility of this model to reject otherwise acceptable model structures.

2. Study Sites

[5] Two basins are studied in these analyses, representing two ends of the hydrological spectrum: humid and semi-arid. The Maimai research catchments are a set of highly responsive, steep, wet, watersheds on the west coast of the South Island of New Zealand. Maimai has a long history of hillslope hydrological research. (see *McGlynn et al.* [2002] for a complete review). The San Jose basin is located in the coastal mountain range of the 8th Region of Chile, a semi-arid area with a Mediterranean climate. It encompasses approximately 750 hectares of hilly terrain that has been intensively cultivated. Streams are generally dry in the summer, while in the winter they respond in a flashy manner to large rainfall events, returning within hours to baseflow levels. Soils in the region are characterized by low conductivities (1–1000 cm/day) and overland flow is observationally the dominant runoff generation process in this region.

3. Methods

3.1. Two Component Hydrograph Separation

[6] Measured runoff from the Maimai catchments has been separated into event and pre-event water components:

$$\frac{Qe}{Qt} = \frac{(Ct - Cp)}{(Ce - Cp)} \tag{1}$$

where C is the $\delta^{18}O$ ‰ of each component, and where subscripts p, *e*, and *t* correspond to pre-event, event and total streamflow, respectively. These data exist for the Maimai catchment from a number of studies [see *McGlynn et al.*, 2002] and average 85–90% pre-event water for storm hydrographs. While these data were unavailable for the San Jose watershed, we estimate that the event water component will be $\gg$50% due to the dominance of overland flow. In each of the following examples, we use very coarse estimates of hydrograph components to establish the value of this type of information for model evaluation.

3.2. Semi-distributed Hydrologic Model

[7] The model used in this analysis is secondary to our stated objective of demonstrating the value of multiple criteria in model evaluation. Nevertheless, a brief description is warranted, as it relates to parameter analysis later in the paper. For a more complete description see *Vaché* [2003]. The approach is semi-distributed conceptual box model where terrestrial model is comprised of a conservation of mass equation:

$$\frac{dV}{dt} = P - ET - K_d - SS_{out} - SOF_{out} \tag{2}$$

where V is the specific volume of water in each element (m), t is current time (days), P is the precipitation rate, ET is the evapotranspiration rate, K_d is the loss to groundwater, SS_{out} is the rate of subsurface outflow and SOF_{out} is the output rate of saturation excess overland flow. An increase in water volume results in an increase in the depth of the saturated zone, and a corresponding decrease in storage of the unsaturated zones. These depths are characterized by model parameters representing soil depth (SD) and porosity (phi). SS_{out} is described as a function of the land surface slope, and the effective conductivity. Effective saturated conductivity (K_s) is assumed to decline exponentially with depth, with the degree of decline modulated by the model parameter m. Infiltration is assumed to occur when the soil is not saturated, but when the saturation deficit reaches zero, infiltration cannot occur. In these instances, excess precipitation is ponded and subsequently delivered directly to the stream network as SOF. Groundwater recharge (Kd) is modeled as a calibrated loss and *a priori* ET estimates are required. An individual mass balance is solved simultaneously for each sub-watershed, and the downslope movement of in-channel water is assumed to proceed kinematically.

3.3. Conservative Tracer: An Additional State Variable

[8] Modeled hydrograph separations are developed using equation (1) in a fashion analogous to that used for the experimental separations. These separations are accomplished through the incorporation of an explicit mass balance of an arbitrary conserved tracer:

$$\frac{dM_t}{dt} = nC_e - tC_t \tag{3}$$

where M_t is the tracer mass within the model unit, n is rainfall rate (m/d), C_e is the concentration of tracer in rainfall, t is flux rate of water out of the model unit and C_t is the concentration of tracer in the model unit (taken from the previous time). C_e is an assumed concentration. The only requirement of this assumption is that it is significantly different from the initial concentration. For single event simulations, a constant C_e and treatment of initial concentrations as representative C_p (from equation (1)) is sufficient to separate the simulated hydrograph. It is important to note that the environmental tracers behind the experimental approach to hydrograph separation at Maimai (and also, in a more conceptual manner, our observations in the San Jose) are affected by zones of immobility and associated flowpath heterogeneity. Like all models, ours is a simplification that does not attempt to include the entire range of heterogeneity found in the real system and there are clear differences then between the model structure and the physical system. Our objective is to evaluate the degree to which those differences affects the

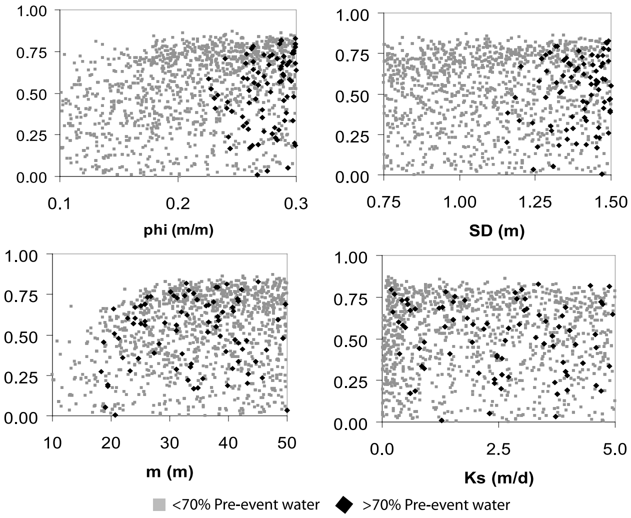

Figure 1. Parameter value versus model efficiency for four model parameters for the Maimai catchment. Models are classified into those simulating <70% pre-event water and those simulating >70% pre-event water.

model partitioning of source waters, defined by discharge and percent old water contribution to the storm hydrograph. If we can establish that the differences are large, we can then successfully reject the model, or parameter sets, as unable to reproduce measurements.

4. Results and Discussion

4.1. Maimai: Parameter Uncertainty Changes

[9] Simulations at Maimai were developed for the period from October 5–October, 21, 1987, with 7 days prior to that used to establish equilibrium between fluxes and volumes within model units. While the maximum model efficiency suggests acceptable model performance (Nash Sutcliffe efficiency of 0.82), the scatter plots (Figure 1) indicate a significant degree of parameter uncertainty. To additionally constrain this uncertainty, we use the modeled hydrograph separation to further discriminate between behavioral and non-behavioral model runs.

[10] As noted previously, calculated values of percent event water have an inherent uncertainty. We cannot, for example, state with precision the percentage of event water in the October storm. But we can strongly argue, based upon the historical isotope separations, that event water comprised less than 30 percent of the peak flow during the period. This threshold is conservative and provides a significant degree of additional understanding of the Monte Carlo results. Table 1 outlines the uncertainty in each model. Set 1 is the set of all models with discharge efficiency greater than 0. Set 2 is further restricted to those simulations with >70 percent pre-event water contributions to stream flow. Uncertainty is calculated as the difference between the 90th and 10th percentiles normalized by the prior parameter range. A larger percentage indicates a larger change, and a positive value indicates a reduction in uncertainty. Parameter uncertainty generally decreases as the definition of a behavioral model is constrained by the more stringent hydrograph separate rules. Parameters indicative of storage – those representing soil depth (SD) and porosity (phi), show the highest level of uncertainty reduction. This suggests that pre-event storage volume estimates are better constrained through the incorporation of hydrograph separation into model evaluation.

Table 1. Uncertainty Results for the Four Parameters Outlined in Figure 1 From the Maimai Simulations

	phi	m	Ks	SD
Set 1	0.78	1.33	4.80	0.73
Set 2	0.20	1.34	4.68	0.22
%Change	74.4	−0.7	2.5	69.9

4.2. San Jose: Overall Model Rejection

[11] Unlike Maimai, very little tracer data exists within the San Jose Basin. Despite the lack of measured hydrograph components, field observational evidence suggests that overland flow is the dominant runoff generation process. The flashy, seasonally ephemeral hydrographs along with a relatively small number of very low hydraulic conductivity measurements corroborate these observations (D. Rupp, personal communication, 2003). Runoff in the San Jose, then, is likely comprised predominantly by event water. However, without the benefit of detailed tracer data we are less certain on the precise split between event and pre-event water. Despite this uncertainty, the coarse assumption that the event water across a storm hydrograph represents greater than 50% of discharge seems reasonable. While infiltration excess overland flow is often cited as the dominant runoff mechanism in semi-arid regions, a growing body of work suggests that saturation excess overland flow may also contribute significantly [*Taha et al.*, 1997], leading us to propose the model outlined above a reasonable hypothesis of runoff generation in the San Jose.

[12] Runoff was simulated over a 3 day period from May 28, 2001 to May 31, 2001, with efficiency reported as a weighted average calculated from three gauging stations. A

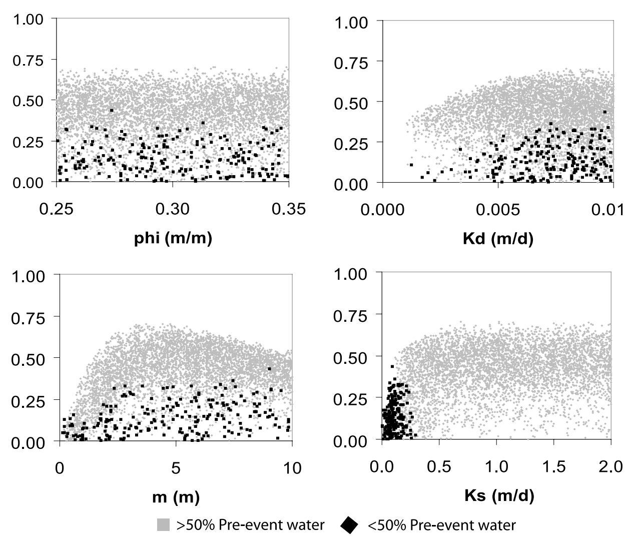

Figure 2. Parameter value versus model efficiency for four model parameters for the San Jose catchment. Models are classified into those simulating <50% pre-event water and those simulating >50% pre-event water.

7 day period was used to establish equilibrium between fluxes and volumes within model units. Scatter plots utilizing a definition of behavioral corresponding to a Nash Sutcliffe efficiency value greater than 0.0, and color classified into greater than and less than 50% pre-event water, are outlined in Figure 2. Maximum discharge based efficiencies approach 0.74, suggesting acceptable model performance. However, the classification indicates that none of the most efficient discharge simulations result in greater than 50% event water contributions. Maximum values of efficiency, in the presence of this additional criteria, fall to approximately 0.43, suggesting that the model (structure and parameters) cannot produce acceptable estimates of discharge *and* an appropriate time source separation.

[13] These results highlight a large set of parameter sets that provide the right answers (based on discharge) but do so for the wrong reasons (based on hydrograph separation). This experience would force us to reject completely this model for application to the San Jose basin. This rejection must lead to alternative model hypotheses and constructs. The first step in this process might be a forensic analysis of why the model failed. Percent event water versus K_s (a basin scale conductivity) (Figure 2) indicates that the value plays a major role in the quantity of simulated event water. Lower values of K_s equate to higher values of event water, consistent with the dominance of overland flow. However, these high event water simulations correspond only to relatively poor simulations of stream discharge. This response appears to be due to the simplified routing procedure for overland flow, where simulated water that can not infiltrate moves to the channel within the time step. A reasonable next step might be to re-evaluate this assumption that overland flow is essentially instantaneous, and include a kinematic wave, time constant, or other mechanism to more effectively route water overland. Perhaps most importantly, we are satisfied to have independently established the inability of this model to correctly simulate flow path dynamics in the San Jose. Thus, soft data, in this case the time source hydrograph composition, is not simply a mechanism to constrain parameter uncertainty, but in fact allows for the overall rejection of a model structure.

5. Conclusions

[14] Our hydrograph separations represent highly uncertain, potentially qualitative statements. At Maimai, we had a reasonable idea of the values based on a significant history of experimental investigations and gauging—but at the San Jose watershed, we relied exclusively on expert definitions of plausible hydrograph compositional percentages. While no intensive monitoring program has been initiated within the San Jose, there is, in fact, a significant amount of *understanding* that can improve *a posteriori* model evaluation. The need to make predictions, and quantify the uncertainty of those predictions in areas with relatively sparse data is a major focus in catchment hydrology today [*Sivapalan et al.*, 2003]. The incorporation of field knowledge and transfer of information between similar basins and runoff regimes into model evaluation has significant utility within this context.

[15] We view the specific benefits of this evaluation as centered in two areas. The first relates to the potential for a more complete understanding of parameter uncertainty. At Maimai, the inclusion of simulated hydrograph composition significantly changed the set of behavioral models and the degree of uncertainty characterizing each parameter. The second focuses on model realism. Conceptual, physically based models are designed to reflect the processes behind the movement of water through catchments. A model that correctly captures discharge and time source composition is more realistic than one that captures only the former. While this realism may not be necessary to adequately describe a discharge hydrograph, it is fundamental to hydro-chemical simulations. Additionally, in some cases a model can perform reasonably well when evaluated for discharge alone, but additional compositional criteria can result in posterior rejection of the model structure itself, as was demonstrated for the San Jose. The incorporation of time source hydrograph separation into evaluation procedures is one mechanism to improve conceptual simulations of catchment hydrology. We also see potential for a host of additional alternative criteria to continue to challenge model structures, applications, and to improve posterior parameter identifiability.

[16] **Acknowledgments.** David Rupp provided data for the San Jose and Jan Seibert was instrumental in the development of our understanding of soft data. Comments from Thorsten Wagner and an anonymous reviewer significantly improved the paper. This research was funded in part by USFS contract number PNW 02-CA-11261930-011.

References

Beven, K. (2001), How far can we go in distributed hydrological modelling?, *Hydrol. Earth Syst. Sci.*, *5*, 1–12.
Burns, D. A. (2002), Stormflow-hydrograph separation based on isotopes: The thrill is gone–What's next?, *Hydrol. Proc.*, *16*(7), 1515–1517.
Buttle, J. M., and J. J. McDonnell (2004), Isotope tracers in catchment hydrology in the humid tropics, in *Forests, Water and People in the Humid Tropics Past, Present and Future Hydrological Research for Integrated Land and Water Management*, edited by M. Bonell and L. A. Bruijnzeel, pp. 750, Cambridge Univ. Press, New York.
Dunn, S. M., C. Soulsby, and A. Lilly (2003), Parameter identification for conceptual modelling using combined behavioral knowledge, *Hydrol. Proc.*, *17*, 329–344.
Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173.
Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*, 2637–2650.
McGlynn, B. L., J. J. McDonnell, and D. D. Brammer (2002), A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, *N. Z. J. Hydrol.*, *257*, 1–26.
Sandstrom, K. (1996), Hydrochemical deciphering of streamflow generation in semi-arid East Africa, *Hydrol. Proc.*, *10*, 703–720.
Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.
Sivapalan, M., et al. (2003), IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*(6), 857–880.
Taha, A., J. M. Gresillon, and B. E. Clothier (1997), Modelling the link between hillslope water movement and stream flow: Application to a small Mediterranean forest watershed, *J. Hydrol.*, *203*, 11–20.
Vaché, K. B. (2003), Model assessment of the effects of land use change on hydrologic response, Ph.D. dissertation, Dept. of Bioengineering, Oregon State Univ., Corvallis, Oreg.
Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Proc.*, *17*(2), 455–476.

J. Bolte, Department of Bioengineering, Oregon State University, Corvallis, OR 97331, USA.
J. J. McDonnell and K. B. Vaché, Department of Forest Engineering, Oregon State University, Corvallis, OR 97331, USA. (kellie.vache@oregonstate.edu)