Predictions in Ungauged Basins: Promise and Progress (Proceedings of symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005). IAHS Publ. 303, 2006.

Process-based strategies for model structural improvement and reduction of model prediction uncertainty

KELLIE B. VACHÉ & JEFFREY J. MCDONNELL

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

kellie.b.vache@agrar.uni-giessen.de

Abstract This paper examines how the combination of simulated discharge and mean catchment residence time (MRT) may be used to subsume flow path process complexity and provide a simple, scalable evaluative data source for water quantity-quality based conceptual models at the catchment scale. A simple Monte Carlo framework is used to evaluate the identifiability of parameters, and how values of mean residence time contribute to the evaluative process and ultimate level of model complexity warranted in the model structure. Our results show that models that might otherwise be acceptable for flow may be wholly rejected for an inability to capture residence time dynamics. The incorporation of soft, or highly uncertain and potentially qualitative data in model evaluation is a useful rejectionist-based mechanism to bring experimental evidence into the process of model evaluation and selection. This may provide a way to reconcile hillslope complexity with catchment scale simplicity and to help define the degree of process complexity needed in a given model application.

Key words modelling; tracers; uncertainty

INTRODUCTION

Increasingly, runoff models form the basis for simulations that address complex environmental problems concerning surface water acidification, soil erosion, pollutant leaching and possible consequences of land-use or climatic changes. Realistic simulations of internal catchment processes related to runoff age, origin and pathway are essential components of these simulations. This paper builds upon recent ideas of Beven (2001, 2002a,b) and brings together flow and streamwater mean residence time (MRT) as complementary evaluation criteria for simple models of catchment runoff that include water quality sensitive flow paths. We use experimentally determined MRT as mechanism to capture the transport time of the catchment model and utilize observational estimates of MRT as *a posteriori* model calibration criteria. The objective of the paper is to use MRT as an additional measure to reduce uncertainty, and guide decisions on the degree of complexity warranted in a rainfall–runoff model. The approach is motivated by Hooper (2001) who advocated better use of hypothesis testing in hydrological modelling. In this regard, we propose a suite of plausible model structures, starting with the most basic configurations for flow and transport. We allow for the rejection of these model hypotheses, and use rejection as a basis for the inclusion of further model complexity.

STUDY SITE

The Maimai research catchments are a set of highly responsive, steep, wet, watersheds on the forested west coast of the South Island of New Zealand. Maimai has a long history of hillslope hydrological research (McGlynn *et al*., 2002). More importantly, unlike other sites where we have done experimental work, Maimai shows striking simplicity in catchment response. The simplicity in catchment response is determined largely by the lack of seasonality and chronically wet state of the system. Soils rarely drain below 90% of saturation and overlie effectively impermeable compacted and cemented conglomerate. Quickflow comprises 65% of the mean annual runoff and 39% of annual total rainfall (*P*) (Pearce *et al*., 1986). Pearce *et al*. (1986) conducted experiments to determine stream water residence time at Maimai. Their work used environmental tracers (i.e. ^2H) of input (rainfall) and output (discharge) to estimate the residence time (using standard techniques described in Maloszewski & Zuber, 1996). Pearce *et al*. (1986) report values of 4 months for the M6 catchment. We argue in this paper that knowledge of MRT is a very useful soft measure of transport, complementary to discharge data normally used for model calibration and evaluation.

HYDROLOGICAL MODEL

We evaluated a set of three models designed to correspond closely to the dominant runoff generation processes at Maimai. Our framework incorporates a variety of different model structures, which we evaluate to find a balance between the need to reduce model complexity, and yet adequately capture the process complexity based on our evaluative measures. While the details behind each of the models is certainly relevant, this short paper is designed simply to outline the utility of MRT to differentiate between models which otherwise appear indistinguishable. For this reason, only a short description is included here. A complete description of the models can be found in Vache & McDonnell (in press). In all cases, Dupuit assumptions are invoked to develop multi-dimensional downslope flow model, following closely from Wigmosta *et al*. (1994). Additionally, the spatial discretization of each model is equivalent—a 10 m grid representing the 3.2 ha catchment, and parameter values are treated as spatially homogeneous. We selected this gridded approach to provide an explicit mechanism to incorporate transient subsurface flows and the land surface slopes that play an important role in driving lateral flow and ageing at Maimai. Three model structures are tested, ranging in number of tuned parameters from 3 to 6. Model 1 (3 parameters) includes a saturated zone but does not consider effective porosity or explicit formulation of an unsaturated zone. Model 2 (4 parameters) includes a saturated zone and an effective porosity term but no explicit formulation of an unsaturated zone. Finally, Model 3 (6 parameters) includes a saturated zone, effective porosity and an explicit formulation of an unsaturated zone. Each of the independent model structures are evaluated under a Monte Carlo framework using a uniform distribution to randomly sampled parameter values from within a prior distribution designed to encompass the range of potential parameter values.

MRT of each simulation was estimated following Goode (1996) using a simulated conservative tracer model with an application as a spatially uniform impulse injection of tracer. Given this tracer application, the MRT can be defined as:

$$MRT = \frac{\int_0^\infty tc\,\mathrm{d}t}{\int_0^\infty c\,\mathrm{d}t} \qquad (1)$$

where *c* is breakthrough concentration and *t* is time. Strictly speaking, this MRT is equivalent to that defined through convolution (the approach used by Pearce *et al.* 1986 based on field collected isotope data) only when the direct simulation incorporates the same flow path distribution as is incorporated by the isotope-based procedure. Of course, the model is a simplification—by design it does not incorporate the full catchment heterogeneity—and so it may not be reasonable to anticipate equivalent MRT. Nevertheless, our goal is to evaluate the degree to which the simplification affects model results and if we can establish that the MRT differences are large, we can then reject the model and use that as a sound basis to incorporate (iteratively) additional complexity.

We evaluated the models using stream discharge and stream water residence times at the outlet of the M8 catchment. Discharge efficiency is defined as the Nash-Sutcliffe measure (R_{eff}). Streamwater MRT is represented as scalar quantity, and is therefore reported as a percentage of measurement value.

RESULTS

Input data at Maimai were available from 3 September 1987 to 30 December 1987 (Fig. 1). The three models were run under a Monte Carlo framework at a time step of 0.01 days, with results collected every 0.2 days. The first 30 days of simulation were allocated towards the stable redistribution of the initial conditions, with model efficiencies calculated based upon results after that point in time. Tracer application occurred at the first time step on the 30th day of simulation.

Each of the three models was run 2000 times, with parameter data, efficiencies and MRT collected for simulations with Nash-Sutcliffe discharge efficiencies over 0.0. Time series data representing modelled discharge for those simulations over 0.75 discharge efficiency were also collected. All models were essentially equally capable of simulating discharge dynamics (Fig. 1). However, MRT values varied considerably between these models (Fig. 2). For models 1 and 2, the simulations with $R_{eff} > 0.75$ all resulted in consistently short MRT values. However, MRT for the set of high R_{eff} in the third model were considerably more variable, with many simulations resulting in MRT which closely approximate the measured value of 120 days. These relationships are further outlined through an analysis of parameter space (Fig. 3).

Model 1

Model 1 was able to reproduce the discharge response of the catchment, with maximum R_{eff} values of over 0.85, however none of the parameter vectors resulted in

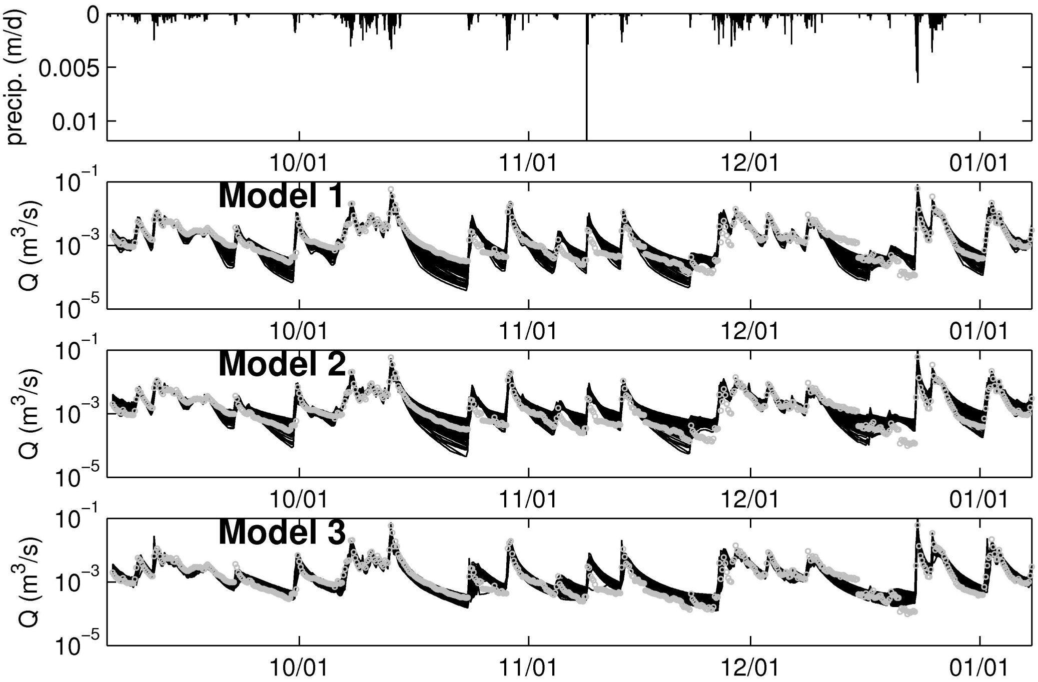

Fig. 1 Simulation results for stream discharge. The Y axis is log transformed to outline more clearly model results at lower discharge. The calibration strategy focused on untransformed R_{eff}, and peak flows are correspondingly better captured. The plotted simulations are those found with Nash-Sutcliffe efficiencies over 0.75. Measured values are plotted as crosses.

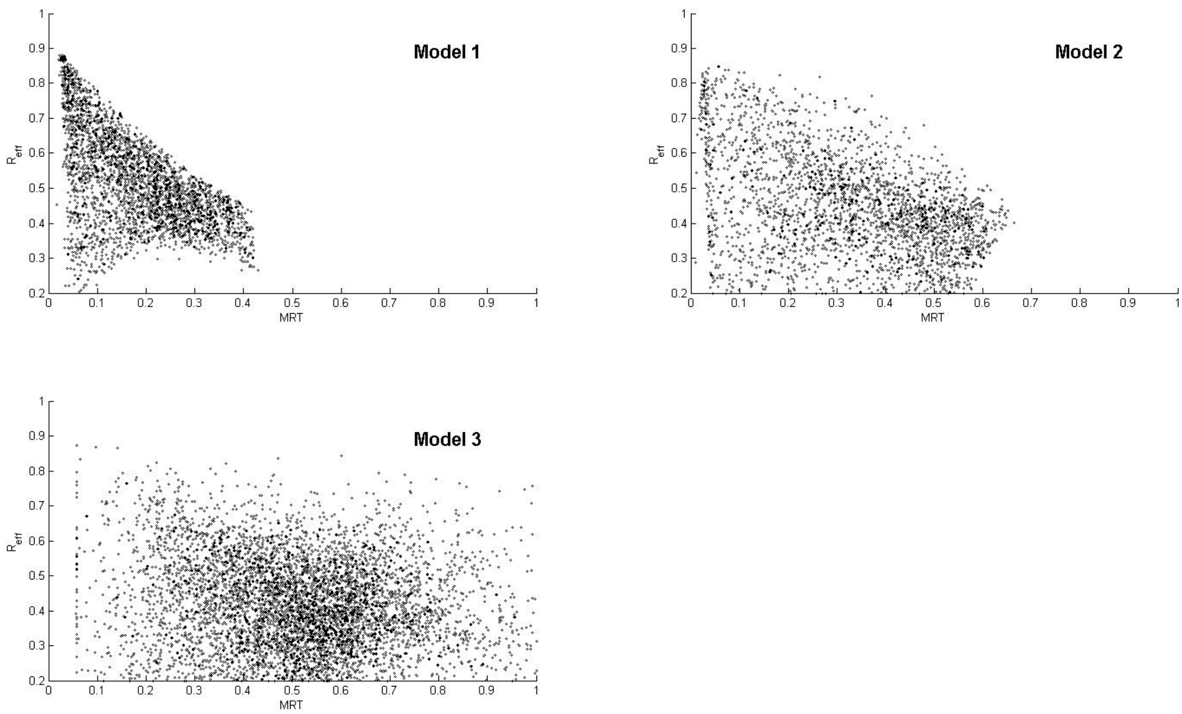

Fig. 2 Scatter plots of discharge versus MRT efficiencies for the 3 models. Discharge efficiency is the Nash-Sutcliffe statistic and MRT efficiency is reported as the ratio of the modelled value to the measured value (120 days). Some values of MRT efficiency were greater than 1, and the bottom axis is truncated for clarity.

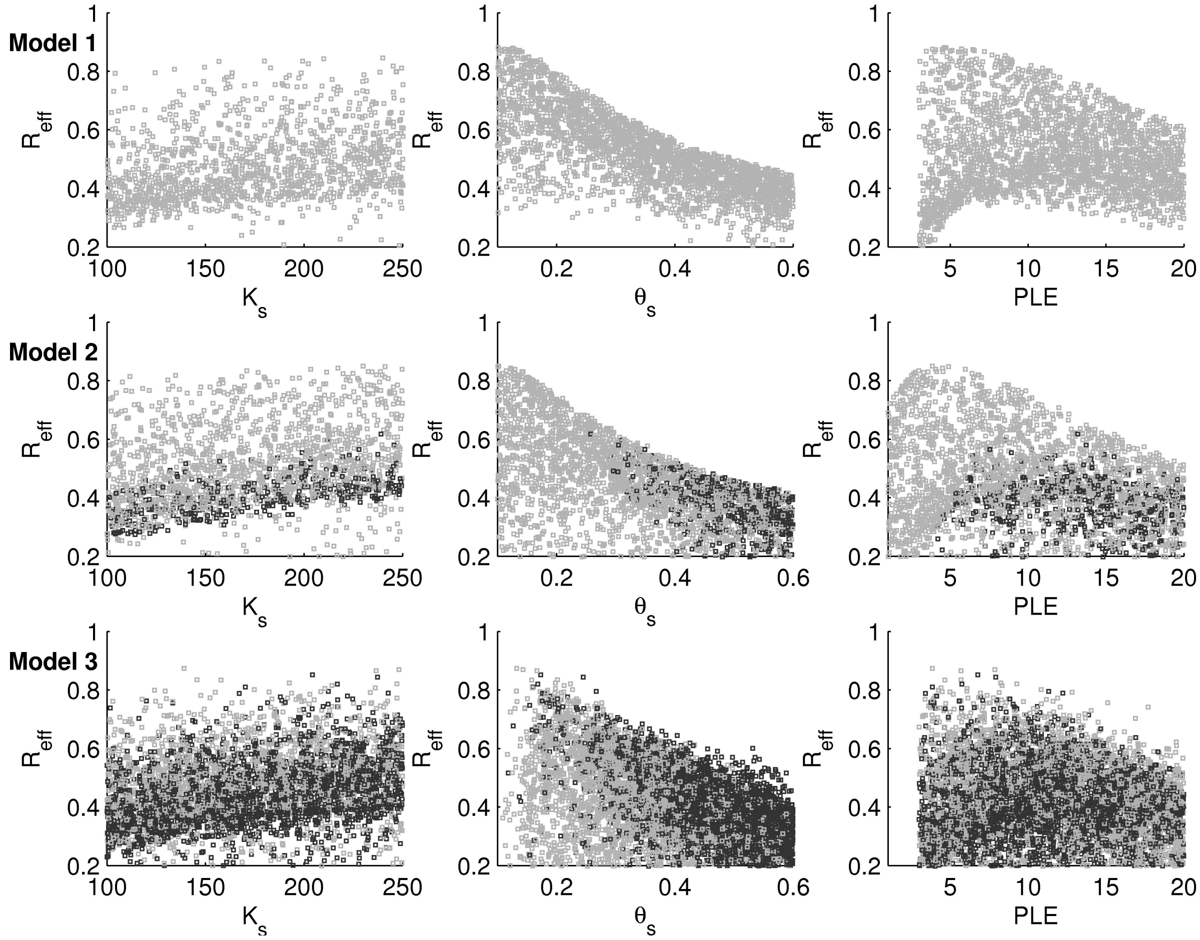

Fig. 3 Scatter plots representing the parameter space for the three parameters included in all of the models. Red models have MRT within 50% of measured. Model 3 is the only structure that results in a set of models with both high discharge efficiency and MRT within 50% of measurements. K_s represents saturated hydraulic conductivity (m d^{-1}), θ_s represents the porosity, and PLE is a power law exponent which modulates the decline in transmissivity with water table height.

stream MRT within even 50% of the measured values. If discharge alone was the evaluative criteria, this relatively parsimonious model with its effective simulations of discharge, might be considered an adequate model structure. But the inclusion of estimates of MRT as evaluative criteria provides an additional perspective on the transport component of the model. Unlike the wave celerity, the transport behaviour is fundamental to solute disposition in the catchment. The fact that MRT across the prior parameter range did not approximate either the magnitude or range of isotopically measured MRT indicates clearly that the three-parameter model, while acceptable for discharge, did not effectively simulate MRT.

Model 2

We hypothesized that model 2 would likely result in longer MRT values than model 1, due to a distinction between dynamic catchment volume responsible for the discharge response and more restrictive volume available in the tracer response with the added parameter (as incorporated with a tracer specific volume parameter). While this model can produce longer MRT, maximum values of simulated MRT were within only ~65% of the measurement values, and the longest MRT values occurred only for relatively poor simulations of discharge (Fig. 2). A clear tradeoff exists between discharge efficiency and MRT for model 2.

The inability of both models 1 and 2 to acceptably capture the stream MRT in the catchment (within a single set of parameters), lead to the conclusion that both models are an overly simplistic representation of catchment processes. Given this result it is clear that within this general gridded model structure, additional model complexity (and therefore parameters) is necessary to capture the MRT and discharge dynamics of the catchment.

Model 3

The addition of an explicit representation of the unsaturated zone dynamics, in combination with the effective tracer volume, results in a model where discharge and MRT are both acceptably reproduced, for some parameter combinations (Fig. 2). This result supports the argument that the complexity involved in the inclusion of both the explicit unsaturated zone and a mechanism to differentiate between dynamic and total storage represents a model structure that successfully reproduces discharge dynamics and stream MRT.

Parameter uncertainty

The incorporation of multiple evaluative criteria provide significant constraints on *a posteriori* model uncertainty, in addition to the potential to reject otherwise acceptable model structures. Model 1 was unable to successfully reproduce stream water MRT, leading us to reject it entirely. Model 2 however, did successfully reproduce both discharge and MRT, albeit with a significant tradeoff in efficiency between the two. In the case that poorer values of discharge efficiency are acceptable, there is a set of model parameter vectors that does in fact produce long MRT and discharge efficiencies over 0.2. While relatively poor in terms of discharge efficiency, this level of discharge reproduction may in some cases be acceptable. In this case, the additional criteria significantly constrain the posterior parameter distribution for some of the parameter values. The simple binary threshold utilized in Fig. 3 indicates that the total porosity is significantly constrained using these additional criteria, with larger values tending to result in improved estimates of MRT. Similar arguments can be made based upon results from model 3.

DISCUSSION

Multi-criteria model calibration studies have been completed in recent years involving saturated area mapping (Franks *et al.*, 1996), groundwater levels (Kuczera &

Mroczkowski, 1998), more complete exploitation of the information content in discharge (Boyle *et al*., 2000), time source components of storm flow (Seibert & McDonnell, 2002; Vaché *et al*., 2004), and geographic source components (Scanlon *et al.*, 2000) as integrative multi criteria tools. However, the development of evaluation criteria comparable to that of a discharge—that are both integrative and scalable—has remained elusive. Also, relatively little guidance has been given in the literature to date on what measures might best constrain a realistic simulation of flow and transport. We argue in this paper that the combination of flow and MRT provide a meaningful test of hydrological models. A small set of models, each progressively more complicated, was tested against these observations. While many of the decisions regarding these structures and their acceptability were subjective, they were designed with a perceptual model and process scale of interest in mind. It is not entirely surprising that the more highly parameterized models were the most successful—the increased degrees of freedom and effective mixing parameter suggest a wider range of potential results. However, our tuned parameter numbers were modest compared with many hydrological models. More important is the fact that simpler models could not reproduce observations, and that we can identify clearly the point at which an increasing degree of complexity was successful (or not). In this case, model 3 was a successful stopping point because it met our initial criteria: the perceptual model and scale of interest were captured, and the discharge and MRT were reproduced to within our criteria of acceptability.

CONCLUSIONS

Improvements in model structures and reduction of model prediction uncertainty will come as we devise new ways to capture process detail into more integrative measures (Sivapalan, 2003). We argue in this paper that MRT may provide one such measure of flow path heterogeneity useful in model structural evaluation and testing. Since conceptual, physically based models are designed to reflect, with varying degrees of complexity, the main stocks and flows of water through catchments, a model that correctly captures discharge and MRT is more realistic than one that captures only the former. More importantly, in some cases a model can perform reasonably well when evaluated for discharge alone, but additional criteria can result in rejection of the model structure itself, as was demonstrated for the most of the model structures we evaluated at Maimai. The incorporation of MRT into evaluation procedures is one mechanism to help understand the limitations of conceptual simulations with water quality sensitive flow paths, and to independently assess the need to incorporate additional process detail or heterogeneity.

REFERENCES

Beven, K. J. (2001) How far can we go in distributed hydrological modelling? *Hydrol. Earth. Syst. Sci*. **5**(1), 1–12.

Beven, K. J. (2002a) Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. *Hydrol. Process.* **16**(2), 189–206.

Beven, K. J (2002b) Towards a coherent philosophy for environmental modelling. *Proc. Roy. Soc. Lond.* **458**, 2465–2484.

Boyle, D. P., Gupta, H. V. & Sorooshian, S. (2000) Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour. Res.* **36**, 3663–3674.

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

Goode, D. J. (1996) Direct simulation of groundwater age. *Water Resour. Res.* **322**, 289–296.

Hooper, R. P. (2001) Applying the scientific method to small catchment studies: a review of the Panola Mountain experience. *Hydrol. Processes* **15**, 2039–2050.

Kuczera, G. & Mroczkowski, M. (1998) Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resour. Res.* **34**, 1481–1489.

Maloszewski, P. & Zuber, A. (1996) Lumped parameter models for the interpretation of environmental tracer data. *Manual on Mathematical Models in Isotope Hydrogeology,* TECDOC-910, 9–58. International Atomic Energy Agency, Vienna, Austria.

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

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

Scanlon, T. M., Raffensperger, J. P. & Hornberger, G. M. (2001) Modelling transport of dissolved silica in a forested headwater catchment: Implications for defining the hydrochemical response of observed flow pathways. *Water Resour. Res.* **37**, 1071–1082.

Seibert, J., Rodhe, A. & Bishop, K. (2003) Simulating interactions between saturated and unsaturated storage in a conceptual runoff model. *Hydrol. Processes* **172**, 379–390.

Seibert, J. & McDonnell, J. J. (2002) On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resour. Res.* **3811**, 1241.

Sivapalan, M. (2003) Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection? *Hydrol. Processes* **175**, 1037–1041.

Vaché, K. B. & McDonnell, J. J. (2006) A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resour. Res.* (in press).

Vaché, K. B., McDonnell, J. J.& Bolte, J. P. (2004) On the use of multiple criteria for *a posteriori* parameter estimation, *Geophys. Res. Lett.* **31**, L21504, doi:10.1029/2004GL021577.

Wigmosta, M. S., Vail, L. W. & Lettenmaier D. P. (1994) A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.* **30**, 1665–1679.