Learning from model improvement: On the contribution of complementary data to process understanding

Fabrizio Fenicia,[1,2] Jeffrey J. McDonnell,[2,3] and Hubert H. G. Savenije[2]

[1] A priori determined model structures are common in catchment rainfall-runoff modeling. While this has resulted in many ready-to-use modeling tools, there are several shortcomings of a one-size-fits-all model structure. The uniqueness of catchments with respect to their hydrological behavior and the need to adapt model complexity to data availability challenge this status quo. We present a flexible approach to model development where the model structure is adapted progressively based on catchment characteristics and the data described by the experimentalist. We demonstrate this approach with the Maimai catchment in New Zealand, a location with a large availability of data, including stream discharge, groundwater levels, and stream isotope measurements. Different types of data are introduced progressively, and the architecture of the model is adjusted in a stepwise fashion to better describe the processes suggested by the new data sources. The revised models are developed in a way to strike a balance between model complexity and data availability, by keeping models as simple as possible, but complex enough to explain the dynamics of the data. Our work suggests that (1) discharge data provides information on the dynamics of storage (represented by the "free" water in the reservoirs) subject to pressure wave propagation generated by rainfall into the catchment, (2) groundwater data provides information on thresholds and on the contribution of different portions of the catchment to stream discharge, and (3) isotope data provides information on particle transport and mixing of the rainfall with the storage present in the catchment. Moreover, while groundwater data appear to be correlated with discharge data, and only a marginal improvement could be obtained adding this information to the model development process, isotope data appear to provide an orthogonal view on catchment behavior. This result contributes to understanding the value of data for modeling, which may serve as a guidance in the process of gauging ungauged catchments.

1. Introduction

[2] The problem of evaluating a model with respect to its ability to reproduce multiple criteria and behavior of a real system is a major focus in hydrology today [*Duan*, 2003]. The current view on model evaluation establishes that a proper model application has to be supported by the following requirements [*Wagener*, 2003; *Gupta et al.*, 2005]: (1) the model must be able to reproduce with accuracy and precision the observed system response, (2) model parameters must be well identifiable given the information extracted by the available data, and (3) the model must be a realistic representation of the system, that is, it should be consistent with our understanding of reality.

These requirements summarize findings of previous research, which revealed issues associated with parameter identifiability [e.g., *Duan et al.*, 1992; *Beven and Binley*, 1992], measure of information [e.g., *Gupta et al.*, 1998] and model realism [e.g., *Seibert and McDonnell*, 2002].

[3] Despite these axioms, the dominant practice in conceptual modeling concerns the application of a priori determined model structures. This one-size-fits-all approach does not help with the fulfillment of the requirements for a proper model application. In fact, there is little chance that a model developed for specific climatic, geologic or hydrologic conditions will represent the hydrological processes realistically and suitably in different environments [*Beven*, 2000b]. Moreover, it is unlikely that the model's complexity reflects a proper balance between parameter identifiability and the model's ability to accurately reproduce the observed system response [*Wagener et al.*, 2001]. These considerations have motivated the development of alternative approaches to hydrological modeling. These include the dominant process concept [*Grayson and Blöschl*, 2000], the "top-down" or "downward" approach, originally introduced by *Klemes* [1983] and reformulated by *Sivapalan et*

[1]Public Research Center – Gabriel Lippmann, Belvaux, Luxembourg.
[2]Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands.
[3]On leave from Department of Forest Engineering, Oregon State University, Oregon, USA.
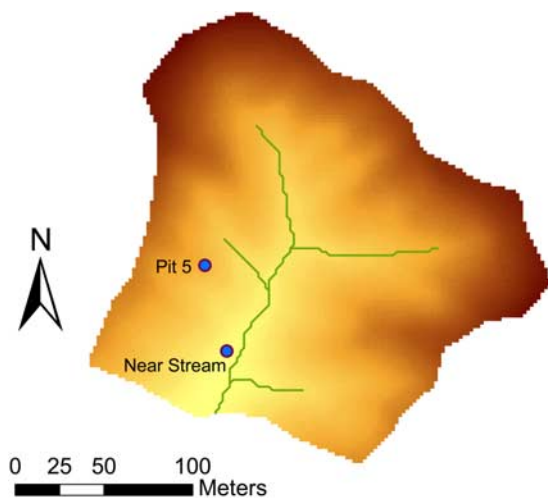
Figure 1. Representation of the Maimai M8 catchment.

al. [2003], the databased mechanistic approach [*Young and Lees*, 1993; *Young*, 2003], the development of flexible modeling frameworks [*Wagener et al.*, 2001] and the balanced qualitative-quantitative approach [*Weiler and McDonnell*, 2007]. These approaches favor the development of models that strike a balance between model complexity and data availability, adapting to the requirements of specific applications.

[4] While much progress has been made in changing the culture of rainfall-runoff modeling, there are still several issues that need to be addressed in order to meet the above mentioned requirements. In the process of conceptual model building, there is in fact little guidance on what dominant model components to include in order to provide a physically meaningful description of a given catchment. Another important aspect concerns the data collection process in relation to modeling. Issues related to the design of hydrologic data networks, and to the value of the resulting information for modeling have long been discussed [*Moss*, 1979a, 1979b; *Langbein*, 1979]. However, at present there is still little guidance for the experimentalist on what to measure, in what order and why, with a view to modeling. As a result, often only a small part of the collected information is used in the modeling process, affecting the quality and level of detail of catchment process representation. These things cloud model realism and hamper understanding of catchment behavior.

[5] This paper presents a new iterative approach in the framework of "scientific investigation", which is a learning process where different aspects of a system are progressively illuminated as the study proceeds [*Box and Tiao*, 1973]. This approach combines elements of the "top-down" philosophy, which consists of deriving catchment internal behavior through an analysis of its response. In line with this philosophy, we start with a simple model structure to reproduce the system response and then progressively refine it based on its shortcomings [*Jothityangkoon et al.*, 2001; *Son and Sivapalan*, 2007]. The analysis of the observed signal and the interpretation of its physical meaning is performed through a discussion between modeler and experimentalist [e.g., *Seibert and McDonnell*, 2002]. This interaction favors the formulation of realistic hypotheses on

catchment behavior. The resulting model can be considered a posteriori determined, in the sense that additional hypotheses on the catchment behavior are progressively incorporated based on the model's failures.

[6] The case study focuses on an application on the Maimai catchment in New Zealand. We use this location to develop a proof-of-concept model structure by introducing progressively different forms of data: first discharge, then groundwater dynamics, and finally stream isotope data. The final model can be considered as a la carte, being necessarily related to the specific characteristics of the selected location. While the developed model is of limited utility as it is related to a specific area, there are elements of this study of more general significance. These are the contribution of our approach to the traditional modeling practice, and our results on the value of complementary data for process understanding. These results are particularly important for the research on ungauged catchments, as they provide general indications on how different types of measurement can contribute to an improved understanding of catchment behavior.

2. Study Area Description

[7] The catchment used for this study is part of the Maimai study area, a set of research catchments in the South Island of New Zealand. These catchments have been the subject of intensive research since 1974 (for review see *McGlynn et al.* [2002]). These study catchments are well suited for this research because of: (1) the availability of complementary data at fine timescales and (2) the strong process understanding already developed at this site which favors the data interpretation in the model development process.

[8] We use data from the M8 catchment, a small forested headwater catchment that drains an area of 3.8 ha (Figure 1) (data available at www.cof.orst.edu/cof/fe/watershd/Documents/Maimai_Data_Note/Maimai.htm). A detailed description of its hydrological characteristics is given by *Pearce et al.* [1986]. The topography of the area is characterized by steep (average 34°) short slopes, with a local relief of about 100 m. The channels are deeply incised, so that the area of near-channel valley-bottom is highly restricted. Soil depth averages at 60 cm, with larger deeper soils in hillslope hollows (topographic convergent zones) and shallower depths near the catchment divide and in the proximity of the stream. Soils are underlain by a firmly compacted and effectively impermeable [*Mosley*, 1979] cemented conglomerate bedrock, which promotes the lateral movement of water in the soil layer above. The overstory vegetation consists of replanted Radiata pine. Annual average rainfall is 2600 mm/a, which produces an average annual discharge of 1550 mm/a, with little seasonal variation. As a result of the humid and wet environment, in conjunction with the catchment characteristics, soil water content rarely declines below 10% of saturation for most of the hydrologic year [*Mosley*, 1979].

[9] The Maimai catchments are among the most hydrologically responsive forested headwater catchments documented in the literature [*Sklash et al.*, 1986]. Soil infiltration capacities exceed by several orders of magnitude typical rainfall rates, hence subsurface flow dominates the production of storm runoff [*Pearce et al.*, 1986]. Analysis of

naturally occurring stable isotope concentrations by *Pearce et al.* [1986] found a fairly stable isotope signal in the stream during rainfall events. From this and other studies it was determined that the contribution of event water to storm runoff was usually less than 25%.

[10] The data used for this study cover a four month period from the beginning of September to the end of December 1987. The data include rainfall, discharge (measured at a 90° V notch), potential evaporation (E_p), groundwater levels, and deuterium measurements in the rainfall and at the catchment outlet. The data timestep used for this study is 20 min. Potential evaporation data were transformed from daily amounts to hourly, using a sine curve distribution between the hours of 6:00 and 18:00 [*Vaché and McDonnell*, 2006]. Groundwater data were available at two locations (Figure 1), one near the stream, and the other at an intermediate location between the catchment divide and the stream (see *Freer et al.* [2004], for a description of these groundwater data). Isotope data covered the whole rainfall period, while discharge isotope data were available only for selected events. A detailed description of the isotope collection approach is given by *McDonnell et al.* [1991].

3. Methodology

3.1. General Framework

[11] Our general methodological framework stems from *Box and Tiao* [1973], where they noted that scientific investigation is a learning process where various aspects of a system are illuminated as the study proceeds. It consists of an iterative process where a tentative model suggests an experiment, and the experiment leads to a modified model. The main aspects that determine the success of this approach are (1) the ''power'' of model diagnosis in demonstrating to what degree the model is representative of the real system behavior [*Wagener and Gupta*, 2005], (2) the efficiency of the experimental design with respect to the modeling purposes, and (3) the value of the inference that can be drawn from the comparison between model and data. This process relies both on the value of established theories which form the ''Science'' foundation of the analysis, and on the creativity, intuition and knowledge of the researcher who applies his or her understanding to the model and experimental work. This represents the ''Art'' aspect of modeling [*Fenicia et al.*, 2008].

[12] Several aspects of this approach are not new to the hydrologic culture. However, the full power of scientific investigation has seldom been expressed in hydrological modeling. In fact, while it has often been reported that conceptual model development should consider a continuous revision of model assumptions [e.g., *Beven*, 2000a], applications that implement this view are few [*Vaché and McDonnell*, 2006]. As a result, model development has rarely been considered in connection to experimental design in a learning process.

[13] The methodology adopted in this paper outlines the various steps of the iterative model development process. We start with a simple model of catchment behavior, which is subsequently tested against observations. We then refine the model structure through an analysis of its shortcomings. In the model development process, we introduce progressively different types of data, at first ''hard'' forms of data

(e.g., discharge), and then ''softer'' (more ad hoc, less continuous, and often requiring some additional interpretation) forms of data (e.g., groundwater measurements and stream water stable isotope composition). The various types of data motivate additional model complexity, demonstrating their contribution to process understanding. We name the evolving model SoftModel, to underline that it is a flexible tool that helps us to explain what we observe. The analysis used for model diagnosis is the multi-objective framework proposed by *Gupta et al.* [1998]. This framework allows the simultaneous evaluation of model performance with respect to different performance indicators. In the present case, we evaluate the agreement of each type of data with model output through a single statistical summary. The model evaluation is therefore performed in a three dimensional objective space.

[14] The type of inference that can be drawn from the comparison between model and data is a crucial aspect of the methodology. In this case the model is developed through an exchange between modeler and experimentalist. The experimentalist shows the modeler which *way to go* when more complexity is required, and the modeler shows the experimentalist which constraints the model requires and where additional data are needed. This dialog benefits both the modeler and the experimentalist.

[15] The methodology presented here builds upon ideas that were introduced recently on model rejection [*Vaché and McDonnell*, 2006], virtual experiments [*Weiler and McDonnell*, 2004], the value of soft data and the modeler-experimentalist dialog [*Seibert and McDonnell*, 2002]. The idea of moving from a simpler to a more complex model structure based on its shortcomings in reproducing catchment behavior is part of the top-down philosophy [*Sivapalan and Young*, 2005]. This philosophy has been utilized to develop models of increasing complexity considering the additional constraints posed by data with decreasing timescale [e.g., *Jothityangkoon et al.*, 2001] or by different types of data [*Son and Sivapalan*, 2007]. In this work, similarly to *Son and Sivapalan* [2007], we identify additional complexity introducing progressively different types of data.

[16] In the framework of the top down approach, modeling starts with a simple conceptualization [*Sivapalan and Young*, 2005]. In the present case we start with a single linear reservoir, arguably the simplest way of reproducing the hydrological behavior of a catchment [*O'Kane*, 2006]. The linear reservoir is common for modeling hydrograph recessions [*Fenicia et al.*, 2006] and has been found useful for simulating hydrograph response at coarse temporal scales [*Jothityangkoon et al.*, 2001]. We use this model because it contains minimum assumptions on catchment behavior, which facilitates the model revision process.

3.2. Model Evaluation

3.2.1. Definition of Objectives

[17] The problem of representing the information contained in the data is a topic of active discussion [*Gupta et al.*, 2008]. Leaving aside the question of how many independent pieces of information can be extracted from a single time series, we evaluated model performance with a single statistical summary for each of the time series used. The selection of model performance measures was performed subjectively, based on considerations of the constraints that data are supposed to impose on model behavior. The

assessment of model performance was also complemented by visual inspection of the simulated and observed time series.

[18] The accuracy of simulation with respect to the observed discharge was evaluated with the objective function F_q, expressed as:

$$F_q = 1 - I_{NS} = \frac{\sum_{k=1}^{n} \left(Q_{o,k} - Q_{s,k} \right)^2}{\sum_{k=1}^{n} \left(Q_{o,k} - \overline{Q}_o \right)^2} \quad (1)$$

where F_q transforms the Nash and Sutcliffe Index, I_{NS}, into a minimization objective, and takes a value of zero for a perfect fit, n is the number of time steps where discharge is evaluated, Q is discharge, the subscripts o and s stand for observed and simulated respectively, overbar indicates an average over the observation period and the index k indicates the specific time step. This performance measure was selected mainly for the purpose of communication, as it is often used in modeling applications.

[19] Model accuracy with respect to the simulation of groundwater was evaluated through the following objective function, related to the correlation coefficient R:

$$F_w^p = 1 - R = 1 - \frac{\sum_{k=1}^{n} \left(S_{s,k}^p - \overline{S}_s^p \right) \left(W_{o,k}^p - \overline{W}_o^p \right)}{\sqrt{\sum_{k=1}^{n} \left(S_{s,k}^p - \overline{S}_s^p \right)^2 \sum_{k=1}^{n} \left(W_{o,k}^p - \overline{W}_o^p \right)^2}} \quad (2)$$

where W represents the groundwater level, S represents the storage level in a model compartment, p represents a certain place or location in the catchment, n is the number of groundwater measurements, and the definition of the other symbols is the same as in equation (1). Groundwater measurements are only available at few locations, and therefore hardly comparable with model variables, representing averages over much larger areas [Beven, 2001]. However, as in the work of Seibert and McDonnell [2002], this objective function was selected to force the model to match the groundwater dynamics, rather than the exact water levels.

[20] In order to evaluate model performance with respect to the stream water deuterium concentration, the following objective function was constructed:

$$F_i = \frac{\sum_{k=1}^{n} \left(\varepsilon_{i,k} \right)^2}{\sum_{k=1}^{n} \left(\delta D_{o,k} - \overline{\delta D}_o \right)^2} \quad (3)$$

$$\varepsilon_{i,k} = \begin{cases} abs\left(\delta D_{o,k} - \delta D_{s,k} \right) - \varepsilon_{\max}, & if \left(abs\left(\delta D_{o,k} - \delta D_{s,k} \right) - \varepsilon_{\max} \right) > 0 \\ 0, & if \left(abs\left(\delta D_{o,k} - \delta D_{s,k} \right) - \varepsilon_{\max} \right) < 0 \end{cases}$$

$$(4)$$

Where, D is deuterium, n is the total number of isotope measurements, ε is the model error expressed by equation (4), $\varepsilon_{\max} = 1‰$ is the standard deviation associated to deuterium measurements, and δD is the relative isotope content of a sample in comparison with the "Vienna Standard Mean Ocean Water" or VSMOW [e.g., Kendall and Caldwell, 1998] represented in parts per thousand (‰). The definition of other symbols is the same as in equation (1). The F_i objective function was constructed to consider as equally good isotope simulations that differ from observations less than $\varepsilon_{\max}$. We emphasize that instead of referring to derived parameters of measured isotopes, such as mean residence time, or fraction of new and old water, which have been adopted as additional criteria for model evaluation [e.g., Vaché and McDonnell, 2006; Son and Sivapalan, 2007], in this case we used the raw isotope data themselves. As mean residence time and new water old water fractions are derived from the isotope signal with very simplistic models, the use of those indices to model evaluation is in principle more uncertain than the direct use of the observations.

3.2.2. Model Calibration

[21] Model calibration was performed both in a single and in a multi-objective framework, depending on the number of objectives to optimize in the different stages of model development. For single objective optimization the parameter search was performed using the PSO (Particle Swarm Optimization) algorithm [Birge, 2003]. The PSO algorithm is a population-based search algorithm based on the simulation of the social behavior of birds within a flock. It is stochastic in nature much like Genetic Algorithms (GA). Its main elements are the maximum number of iterations N_{PSO}, the population size M_{PSO}, which defines the number of model runs at each iteration, the maximum number of iterations to reach consensus that optimization has occurred K_{PSO}, and an error threshold E_{PSO}. Optimization stops either if N_{PSO} has been reached, or when after M_{PSO} function evaluations the objective function varies less than E_{PSO}. Other parameters were kept constant as suggested by default [Birge, 2003].

[22] For multi-objective optimization we used the NSGA II (Nondominated Sorting Genetic Algorithm, Deb et al. [2002]). The NSGA II is a multi objective GA which, with respect to traditional GAs, has the main advantage of not requiring parameter tuning. For the NSGA II the main elements to get the function running are the number of iterations N_{GA}, and the population size which defines the number of model runs at each generation M_{GA}. The total number of function evaluations is therefore N_{GA} times M_{GA}. Both algorithms (which can be downloaded at mathworks.com) represent state of art in optimization theory [Engelbrecht, 2005; Deb et al., 2002].

[23] As complexity in the model structure is introduced in correspondence of model failures, model parameters have often a specific role in the simulation of catchment response. In order to take advantage of this aspect in the calibration process, we made use of the Stepped Calibration Approach [SCA, Fenicia et al., 2007]. This approach associates groups of model parameters to related calibration objectives, and calibrates each group of parameters with respect to each calibration objective in separate stages. The aim of this approach is to ensure that model components perform the tasks for which they are intended, while not compensating for other processes that are poorly represented. Moreover, the approach narrows the parameter
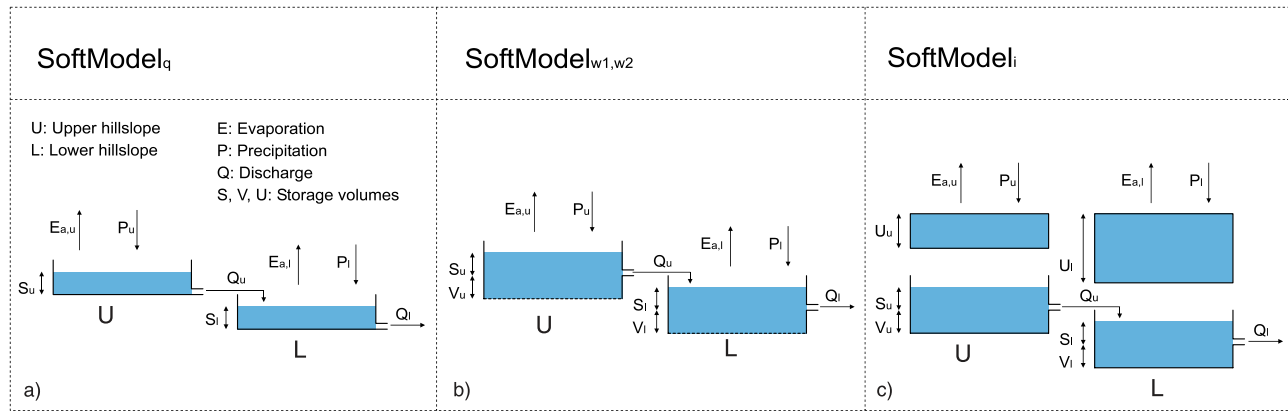
Figure 2. Structure of *SoftModel*$_q$, *SoftModel*$_{w1,w2}$, and *SoftModel*$_i$.

space, and improves the conditions for efficient parameter sampling.

3.2.3. Sensitivity Analysis

[24] An evaluation of parameter sensitivity is important to assess to what extent parameter values are constrained by the data. A lack of sensitivity is a symptom of model over-parameterization, and reduces confidence in the hypotheses describing catchment processes. Approaches to determine parameter sensitivity evaluate how changes in model parameters affect model predictions. Among the different existing approaches [*Tang et al.*, 2007], perhaps the most widely used is Regional Sensitivity Analysis (RSA) [*Hornberger and Spear*, 1981], which has been applied in many variants in different applications [e.g., *Freer et al.*, 1996; *Wagener et al.*, 2003].

[25] In this study, we use elements of the RSA, with the difference that while the RSA is based on uniform sampling of the parameter space, we use the parameter samples generated during model calibration. Hence as the optimization algorithms converge toward the regions of the parameter space that performs best, the analysis is used to express the location and extent of these areas within the feasible parameter space. The spread of the solutions where the sampling concentrates is taken as an indication of parameter sensitivity.

[26] The analysis is applied to the model structure that represents the final stage in the model development process. Like in the RSA, we adopt a partitioning of the parameter space in behavioral (good) and nonbehavioral (bad) solutions, based on the selection of arbitrary thresholds of acceptance. The analysis is performed in a multiobjective way and therefore thresholds were established for each of the three objective functions used during calibration.

[27] For each behavioral parameter set and each objective function, the difference between the acceptance threshold and the objective function value is used to construct a performance measure expressing better performance with higher values [e.g., *Wagener et al.*, 2003]. The performance measures corresponding to each objective function are then rescaled to produce a cumulative sum of unity. The cumulative performance C over the parameter ranges is then used to characterize parameter sensitivity. The total extent of the line representing the cumulative performance indicates the total range where the parameter is identifiable; a straight line indicates poor identifiability within the range,

whereas deviations from a straight line indicate regions of identifiability. Higher gradients indicate a grouping of behavioral solutions, while lower gradients indicate lack of behavioral solutions.

4. Results

4.1. Model Development

4.1.1. Discharge Data; Development of *SoftModel*$_q$

[28] The modeling process that let to development of *SoftModel*$_q$ (Figure 2) started with a single linear reservoir ($Q = S/k$, where k is the reservoir timescale, and S the reservoir storage), widely acknowledged as the starting point for all conceptual rainfall-runoff models [*Eriksson*, 1971]. Actual evaporation was assumed to be equal to the potential evaporation, as the catchment is characterized by little moisture stress and soil moisture remains near saturation for most of the year (the yearly water balance of the catchment also confirms that this is a good approximation [*Mosley*, 1979]). The single parameter representing the reservoir timescale was calibrated by minimizing F_q.

[29] A visual examination of the hydrograph made it clear that the model could not represent the slope of the recession curve over the full range of observed discharge (Figure 3). This motivated the first improvement in the model, which consisted of substituting the linear reservoir for a nonlinear reservoir ($Q = S^{1+\alpha}/k$, where α is a nonlinearity parameter). This modification was preferred over other model options (e.g., introduction of more reservoirs) because it was the most parsimonious in terms of parameters introduced. It in fact resulted in the introduction of only one additional parameter.

[30] The model modification improved the fit substantially (Figure 3). However, there was a lag between observed and simulated time series (where the model produced a peak discharge simultaneously with the peak of the rainfall) that we could not correct with the existing model structure. While several options exist to correct this problem [e.g., the use of a transfer function as in the HBV model of *Bergström*, 1995], we preferred to introduce an additional linear reservoir, in series with the other reservoir [*Nash*, 1960]. The selection of this alternative was motivated by the catchment configuration, characterized by the impermeable bedrock layer that promotes lateral movement of water both during and between events. Experimental work at the site has shown this based on
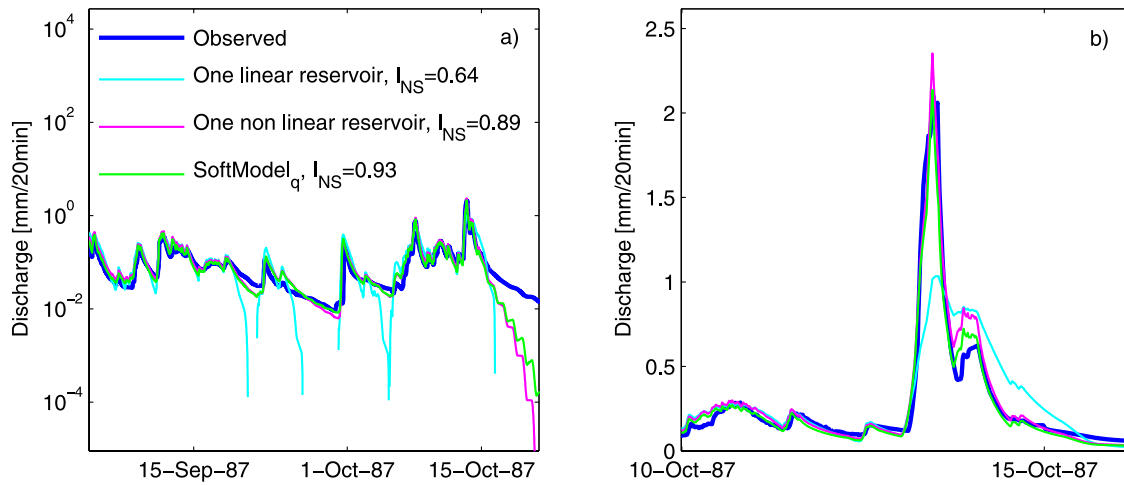
Figure 3. Optimum performance of one linear reservoir, one nonlinear reservoir, and *SoftModel$_q$* with respect to discharge simulation, on a subset of the calibration record.

3D tensiometer monitoring during events [*McDonnell et al.*, 1990] and based on isotopic evidence where water ages clearly in a downslope direction [*Stewart and McDonnell*, 1991]. We represented this feature in the revised model by a series of two cascading reservoirs: a linear upper hillslope reservoir and a nonlinear lower hillslope reservoir (Figure 2a). The linear reservoir is representative of the shallower more uniform soils in the portions of the catchment that are closer to the catchment divide, while the nonlinear reservoir is representative of the thicker soils of topographic convergence zones and accounts for a larger variability of catchment response. The nonlinear reservoir also captures behaviorally the observed decrease of hydraulic conductivity with depth [*Mosley*, 1979]. We describe these simple model decisions by the following equations:

$$Q_u = S_u/K_u \tag{5}$$

$$Q_l = S_l^{1+\alpha}/K_l \tag{6}$$

$$P_u = (1-f)P \tag{7}$$

$$P_l = P - P_u \tag{8}$$

$$E_{a,u} = (1-f)E_p \tag{9}$$

$$E_{a,l} = E_p - E_{a,u} \tag{10}$$

where the subscripts u and l refer to the upper and lower hillslope reservoir, respectively, S (mm) is the reservoir storage, K (t) is the timescale of the reservoirs, P (mm/t) stands for precipitation, E_a (mm/t) and E_p (mm/t) are actual and potential evaporation respectively, and f (−) is the surface fraction of the catchment represented by the lower hillslope reservoir. This model, which we named *SoftModel$_q$*, is characterized four parameters: K_u, K_l, f, and α.

[31] This representation of hydrological processes is simple and not entirely satisfactory for the experimentalist, who is interested in a more mechanistic description of hillslope processes. The process knowledge that has been acquired during the extensive analysis of this catchment suggests a catchment behavior that is more elaborate than this simple structure represents. The horizontal subdivision in two zones (upper hillslope and lower hillslope), and the lumping of the vertical profile in a single unit represent this catchment in a very simplified way. *Seibert and McDonnell* [2002], for example, divide the hillslope in three zones in horizontal direction (hillslope, hollow and riparian), and in two zones in vertical direction (unsaturated and saturated). The lack of a separate description of interception implicitly combines this process into the total evaporation, even if evaporation from water stored on forest canopy accounts for a fraction of 20–30% of gross rainfall [*Pearce et al.*, 1986; *Savenije*, 2004].

[32] However, by looking at the discharge data alone and at the accuracy obtained by *SoftModel$_q$* it is difficult to justify the need for more model complexity (Figure 3). Indeed, the introduction of additional model components would add additional degrees of freedom that would not be supported by the data. The introduction of an interception reservoir, for example, did not produce significant improvements in the final hydrograph, probably as a result of the chronically wet conditions which do not enhance the threshold behavior associated to this process. This however, does not mean that the model definition should not be different, or that it is not possible to extract more information from the discharge data. The model selection process is closely related to a subjective interpretation of the data, while it is known that the same type of data can be employed in more ways. In this case, the satisfactory visual examination of model prediction (Figure 3), the short length of the time series used, and the priority of focusing on different types of information, did not motivate further model development based on discharge time series alone.

4.1.2. Groundwater Data; Development of *SoftModel$_w$*
[33] Groundwater levels were available from two locations in the catchment, one located in a hollow, close to the catchment divide, and another located near the stream. In
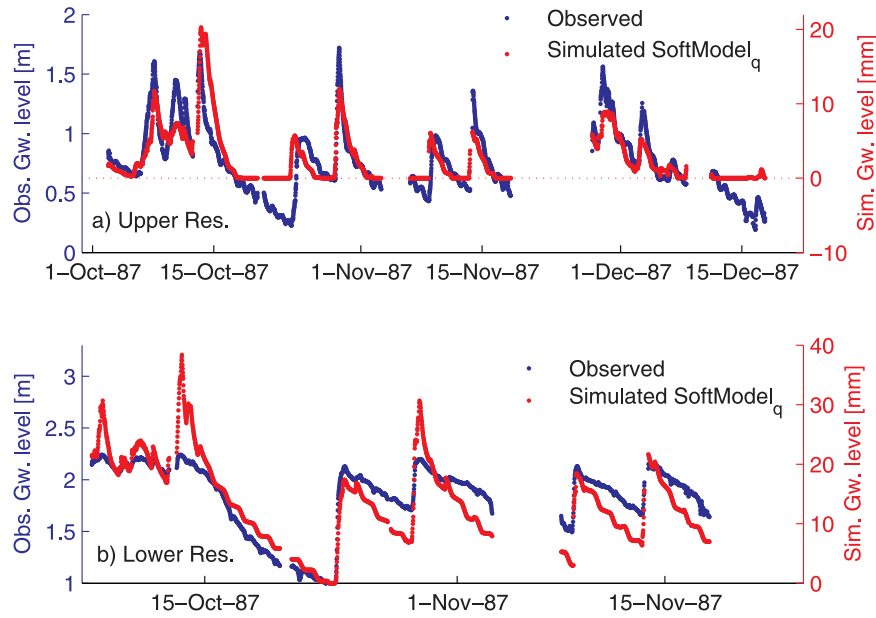
Figure 4. Groundwater level simulation of *SoftModel$_q$*.

order to evaluate model performances at both locations, the measurements from the upper and lower hillslope locations were compared with the levels in the upper and lower reservoir respectively, using a combined objective function:

$$F_w = \frac{1}{2}\left(F_w^u + F_w^l\right) \quad (11)$$

The introduction of an additional criterion puts the model evaluation in a multiobjective environment [*Gupta et al.*, 1998]. The evaluation of *SoftModel$_q$* with respect to both F_q and F_w shows a tradeoff between these objectives, which is described by the Pareto-optimal front (shown later in the paper). In order to evaluate the performance of *SoftModel$_q$* with respect to groundwater visually, the model selected is the one that corresponds to a I_{NS} of 0.9, where the performance with respect to F_w seems to level out. This model is shown in Figure 4. Examination of this graph shows that the overall catchment dynamics are more or less matched, but that storage fluctuations are poorly represented in the upper hillslope groundwater. This may be because the lower limit of reservoir S_u is the zero level of the reservoir. In the lower hillslope reservoir, on the other hand, it appears that the dynamics associated with lower storage levels are better simulated than those associated with higher levels. The lower reservoir apparently has an upper threshold not represented by the model.

[34] These limitations can be used to question the validity of *SoftModel$_q$*, and to develop additional hypotheses of catchment behavior to then be incorporated into the model. A first modification of *SoftModel$_q$* was to allow the reservoir level to go below the threshold of runoff production. This means that below this threshold, additional storage is considered that can be drained by evaporation. This choice is motivated by the consideration that the initiation of runoff in a catchment is a threshold process that occurs when catchment storage rises above a certain level (i.e., the fill and spill hypothesis of *Tromp-van Meerveld and McDonnell*

[2006]). This modification should overcome the limitations displayed in reproducing upper hillslope groundwater dynamics.

[35] The resulting model, named *SoftModel$_{w1}$* (see Figure 2b), contains one additional parameter with respect to *SoftModel$_q$*. This model is defined by equations (5), (6), (7), (8) and by the following equations:

$$E_{a,u} = \begin{cases} (1-f)E_p, & \text{if } S_u \geq 0 \\ (1-f)E_p e^{C_e S_u}, & \text{if } S_u < 0 \end{cases} \quad (12)$$

$$E_{a,l} = \begin{cases} fE_p, & \text{if } S_l \geq 0 \\ fE_p^{C_e S_u}, & \text{if } S_l < 0 \end{cases} \quad (13)$$

where C_e $(-)$ is a parameter that accounts for the reduction of evaporation (in fact in this case it is entirely transpiration) if groundwater levels are below the threshold that activates the reservoirs. As a result of this hypothesis, there is no connectivity between the reservoirs if S_l is below zero, and the reservoirs empty only due to evaporation, as described by equations (12) and (13). At this stage, the lower limit of the reservoir level is left undetermined. The performance of *SoftModel$_{w1}$*, characterized by five parameters, is represented in Figure 5, which demonstrates that model performance improved in the simulation of the upper hillslope reservoir.

[36] To overcome the limitations related to the lower hillslope reservoir, it is proposed that the nonlinear storage-discharge relation is complemented by a nonlinear distribution of storage over the profile. A nonlinear relationship between groundwater level and storage for the lower hillslope was also suggested by *Seibert et al.* [2003] who used a similar model at a different location. In this case, such a relationship is justified by the fact that a significant decrease of drainable porosity with depth was observed, which was associated with different character-
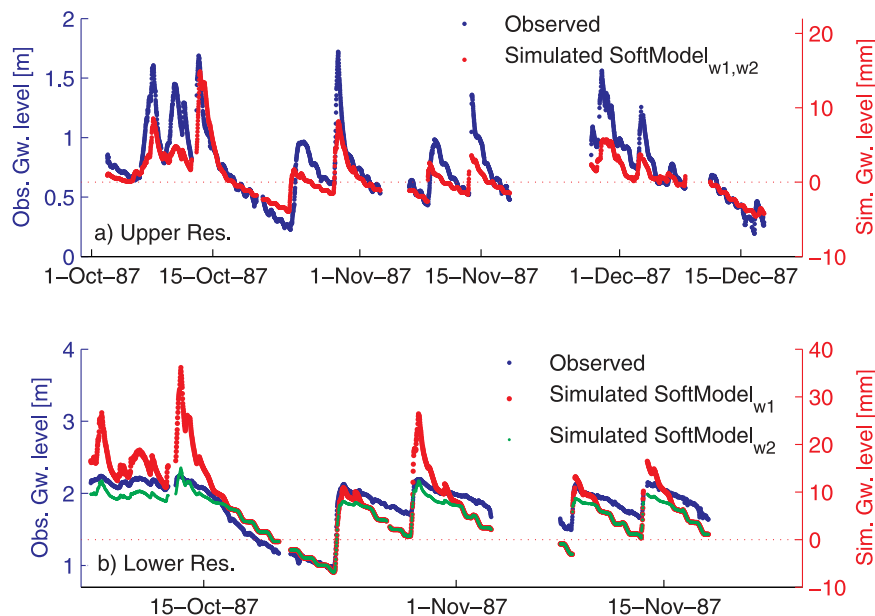
Figure 5. Groundwater level simulation of $SoftModel_{w1}$, and $SoftModel_{w2}$. Model accuracy increases at each model development stage.

istics of the upper organic soil and lower mineral soil [*Mosley*, 1979].

[37] This hypothesis is implemented by introducing a storage threshold for the lower hillslope reservoir S_l^{tr} (mm), and assuming that equal changes of storage above and below this threshold produce unequal changes of groundwater levels by a factor $C_w(-)$. C_w is a number lower than unity that represents the ratio between the drainable porosity in the lower layer and the drainable porosity in the upper layer. The model that includes this modification was named $SoftModel_{w2}$ and is characterized by six parameters. Its performance is also shown in Figure 5. Clearly, this modification only affects the simulation of the water level in the lower reservoir, leaving the simulation of all other fluxes unaltered.

4.1.3. Isotope Data, Development of $SoftModel_i$

[38] Isotope samples were available for rainfall for all events and the major discharge events. Because of the uncertainty in the observed input and output and due to the discontinuities in the measured output, the states of the model prior to each event were updated through state updating. In doing so, the concentration of all model reservoirs prior to each event was set as the concentration that was measured in the base flow prior to the event (i.e., discharge prior to the rainfall event when this measure was available, or as the concentration after the event when low flow conditions were reached if pre-event based was not available). No relative weighing of model output and measured output was performed, and full weight was assigned to the measurements.

[39] The use of state updating avoids that the average of the modeled fluctuations deviates from that of the measurements, and allows the model to fit better the internal dynamics of each event, when it has the potential to do so. The shortcoming is that in updating model states some information about the processes may be lost. However, it

has to be realized that tracer simulations are generally event based. In this case, even with state updating, we try to capture the observed dynamics of the entire observation period within a single model.

[40] The performance of $SoftModel_q$ and that of $SoftModel_{w1}$ and $SoftModel_{w2}$ were tested with the hypothesis of "complete mixing", which assumes that the rainfall reaching each reservoir mixes instantaneously and completely with the available reservoir storage. At this stage, no additional parameters were added and the performance F_i was calculated without recalibrating the model. However, the introduction of the mixing hypothesis required the specification of a mixing volume. Hence a bottom level of the reservoirs (V_u (mm) and V_l (mm)) was selected based on the minimum storage level reached during model calibration (11 and 22 mm respectively).

[41] The behavior of $SoftModel_{w1}$ and $SoftModel_{w2}$ with respect to isotope simulation was the same, since both models differed only by the characterization of the groundwater levels and not by their flows. The performance of the models is shown in Figure 6. Neither of the two models was able to match the observed isotope signal adequately. $SoftModel_{w1,w2}$, characterized by a larger storage, indeed improved the simulation of $SoftModel_q$. This suggests that a better representation of the isotope signal can be obtained by accounting for even larger mixing volumes [*Barnes and Bonell*, 1996]. This hypothesis was implemented in $SoftModel_i$ (represented in Figure 2c). Compared to $SoftModel_{w2}$, the model takes into account an additional storage for each reservoir, of which the volume is identified by the parameters U_u (mm) and U_l (mm). In order to reduce the risk of overparameterization, the ratio of the two is kept constant at a value that is representative for the difference between the soil thickness of the upper and lower hillslope, thus reducing the number of free parameters to one (e.g., U_u). This approach is consistent with the parameterization proce-
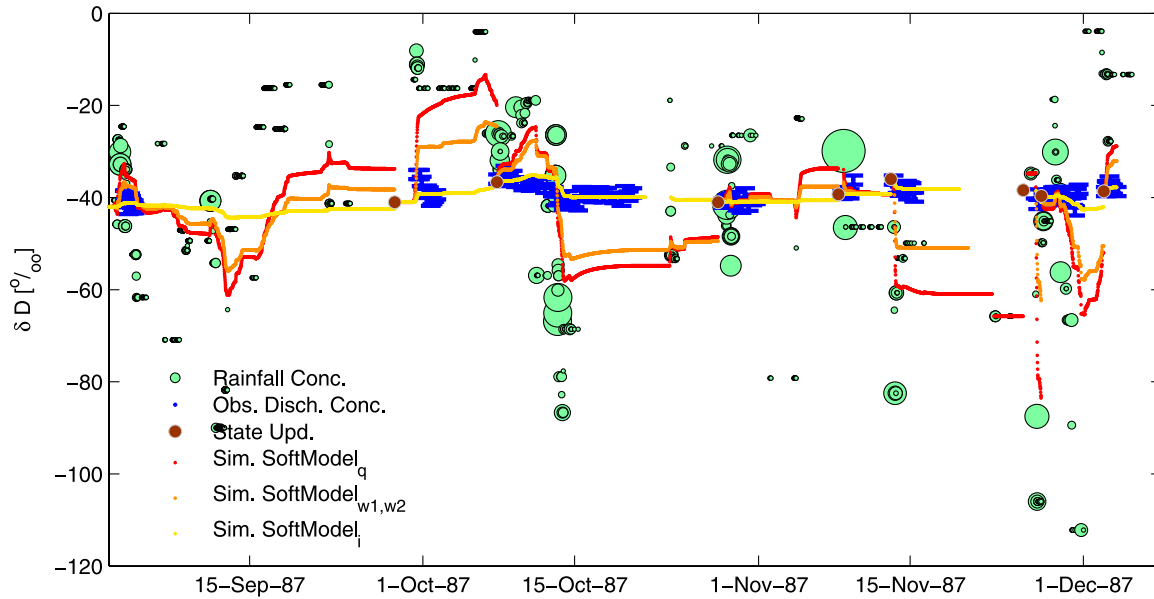
Figure 6. Simulation of Deuterium concentration of $SoftModel_q$, $SoftModel_{w1,w2}$, and $SoftModel_i$.

dure outlined by *Refsgaard and Storm* [1996] who suggest fixing the spatial patterns in order to reduce the number of free parameters that need to be adjusted during calibration. On the basis of the data given by *Seibert and McDonnell* [2002], a ratio U_l/U_u of 3 was selected.

[42] This additional storage accounts for the mixing of rainfall with a catchment storage which is much larger than what can be determined using discharge or groundwater data only. It can be identified with the storage in the unsaturated zone, or with other storages that participate to the rainfall mixing process but cannot freely drain when subject to gravity alone. Hence we assume that the storage is inactive with respect to the ability of the catchment of producing runoff, and that participates only in the mixing process. Moreover, we assume that the storage is fixed in volume, also based on the consideration that moisture state of the catchment does not vary greatly throughout the year. As a result, $SoftModel_i$ is characterized by the same equations as $SoftModel_{w2}$ with respect to flow and groundwater dynamics, while the rainfall is supposed to mix completely and instantaneously both in the total volume of the upper hillslope fraction (i.e., $U_u + S_u + V_u$) and in the total volume of the lower hillslope fraction (i.e., $U_l + S_l + V_l$).

[43] The performance of $SoftModel_i$, characterized by seven parameters, is shown in Figure 6, where it is possible to observe that the overall simulation with respect to isotope data is improved significantly. While the M8 catchment is known to be a well mixed system [*Pearce et al.*, 1986], the hypothesis of complete mixing is clearly a simplification. *Stewart and McDonnell* [1991] show that the age of water increases downslope and vertically in the profile, and *McDonnell et al.* [1991] concludes that between storms mixing is particularly important. We formulated additional hypotheses which go beyond complete mixing, but they did not lead to qualitative improvements in the model performance. While they might have important consequences on the interpretation of the catchment behavior, they go beyond

the philosophy of this work which is to introduce complexity where supported by the data.

4.2. Model Calibration

[44] The model development process favors the connection of model parameters with specific calibration objectives. In order to take advantage of this aspect, the model was calibrated through the SCA. The approach consists in associating groups of parameters with specific objective functions and in calibrating each group of parameters with respect to each objective function in separate stages. An ideal condition for the application of this approach is when model parameters related to a certain objective function are uncorrelated to objective functions associated with preceding calibration stages. This allows the calibration of each group of parameters not to be affected by subsequent calibration steps [*Fenicia et al.*, 2007]. In the present case, by looking at model equations, it is possible to observe that that while all parameters of $SoftModel_q$ and $SoftModel_{w1}$ are correlated to all objective functions, the parameters introduced in $SoftModel_{w2}$ (i.e., S_l^{tr} and C_w) affect only the simulation of groundwater and therefore do not influence the simulation of discharge, and the parameter introduced to allow for isotope simulation (i.e., U_u) in $SoftModel_i$ affects only the simulation of isotope concentration, and not the simulation of groundwater levels and discharge.

[45] Hence while the parameters associated to $SoftModel_q$ and $SoftModel_{w1}$ were calibrated with respect to both F_q and F_w, the additional parameters that characterize $SoftModel_{w2}$ were calibrated with respect to the single objective F_w, and the parameter introduced in $SoftModel_i$ was calibrated with respect to the single objective F_i. In the calibration of $SoftModel_{w2}$ and $SoftModel_i$, the already calibrated parameters were kept at constant values.

[46] The multiobjective optimization of $SoftModel_q$ and $SoftModel_{w1}$ was performed with the algorithm NSGA II, with parameters $N_{GA} = 200$ and $M_{GA} = 100$. The initial parameter ranges defining the feasible parameter space are

Table 1. Initial Parameter Range for Model Calibration

Parameter	Units	Initial Range
K_u	t (20 min)	$4-10^6$
a	...	$0.5-4$
K_l	t (20 min)	$4-10^6$
f	...	$0.1-0.9$
C_e	mm^{-1}	$0.001-3$
S_l^{tr}	mm	$2-50$
C_w	...	$10^{-4}-1$
U_u	mm	$80-110$

reported in Table 1. The samples generated by the algorithm in the objective space are shown in Figure 7, which demonstrates the tradeoff in performance between F_q and F_w and the general improvement associated to the model modification. The performance tradeoff shows that an increase in performance with respect to F_w is at the expense of a performance decrease with respect to F_q.

[47] The single-objective calibration stages of *SoftModel*$_{w2}$ and *SoftModel*$_i$, with respect to F_w and F_i respectively are performed using the PSO algorithm, with parameters $N_{PSO} = 100$, $M_{PSO} = 24$, $K_{PSO} = 35$ and $E_{PSO} = 10^{-4}$. The repetition of these stages for a group of parameter sets (i.e., those of *SoftModel*$_{w1}$ that perform better than 0.3 for F_q, 0.15 for F_w) gives an indication of the general model improvement. In Figure 7 we see that *SoftModel*$_{w2}$ outperforms *SoftModel*$_{w1}$ with respect to F_w, while the performance of the two models is the same with respect to F_q. Figure 8 shows the performance of all models in the three-dimensional objective space. We can see that each improvement causes a shift toward the origin of the axes. Moreover, the increase in performances between *SoftModel*$_{w1}$ and *SoftModel*$_{w2}$ and between *SoftModel*$_{w2}$ and *SoftModel*$_i$ is parallel to the axes F_w and F_i respectively, because the parameters introduced are correlated to one objective function only.

[48] The model outcomes shown in previous figures correspond to selected models from the calibration results. Specifically, Figure 3, which shows the *SoftModel*$_q$ discharge prediction, corresponds to the model that optimizes the single objective F_q. Figures 4, 5, and 6, which represent model performance with respect to groundwater and isotopes, correspond to the optimal models associated to a fixed value of F_q of 0.1 (a I_{NS} index of 0.9).

4.3. Parameter Sensitivity

[49] While the focus of this work is specifically on the contribution of complementary data for model development, an assessment of parameter identifiability is important to evaluate to what extent our approach helps constraining model's degrees of freedom. Parameter identifiability with respect to their associated objective functions (i.e., the objective functions used for their calibration) is expressed in Figure 9. The performance measures based on which the cumulative performance C is calculated were determined from the values of F_q, F_w and F_i, using the respective acceptance thresholds of 0.15, 0.3, and the maximum of F_i. The results of the sensitivity analysis are clearly influenced by the ability of the selected objective functions of enhancing the role of model parameters in the reproduction of the observed system behavior. Model parameters have different

effects on the simulation and not all of them may be well represented by the objective functions used. However, Figure 9 shows that model parameters are well identifiable. The initial parameter ranges are reported in Table 1. The final sensitivity ranges are much narrower than the initial ones, and lie within one order of magnitude for all parameter values, and for most of them within a 10% of their values. It is interesting to notice that parameter identifiability may change with respect to different objective functions. As an example, the parameter K_l is better constrained by F_w than by F_q. In general, the results of the analysis support the conclusion that the iterative approach to model development applied in this study favors the development of models with a degree of complexity that is supported by the data.

5. Discussion

5.1. Adaptive Approach to Model Development

[50] We used information about stream discharge, groundwater level, and isotope concentration to generate a model structure that evolves progressively as each new type of data is introduced. The model incorporates an increasing number of hypotheses that allow evaluating the contribution of different sources of data in representing the catchment behavior. *SoftModel*$_q$, which characterizes the catchment description with respect to discharge data alone, represents the catchment as a series of reservoirs: an upper linear reservoir, and a lower nonlinear reservoir. The description of the catchment structure as a series of reservoirs is motivated by the consideration that the catchment is characterized by an impermeable bedrock layer, which impedes deep infiltration, promoting lateral movement of water. The reason why the storage discharge relation is linear in the upper reservoir and nonlinear in the lower one, is that soils are shallower and more uniform in the upper part of the basin than in the lower one, where they cause a larger variability in runoff response. *SoftModel*$_{w2}$, which describes the catchment behavior with respect to discharge and groundwater data, involves the additional hypothesis that
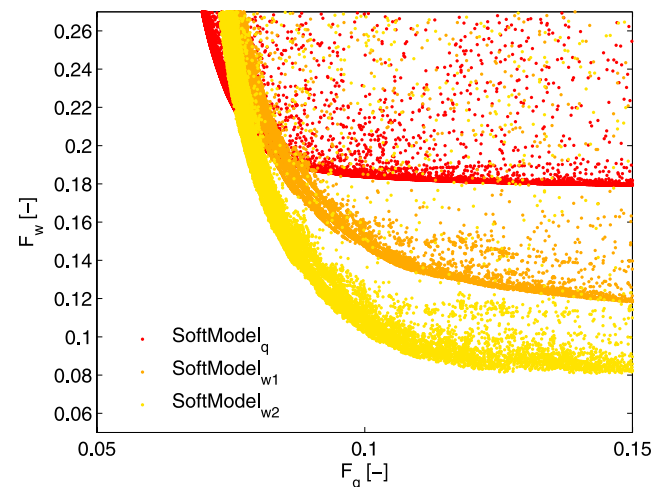

Figure 7. Performance of *SoftModel*$_q$, *SoftModel*$_{w1}$ and *SoftModel*$_{w2}$ with respect to discharge (F_q) and groundwater level simulation (F_w). In model evolution, the performance F_w improves, while F_q remains at the same level.
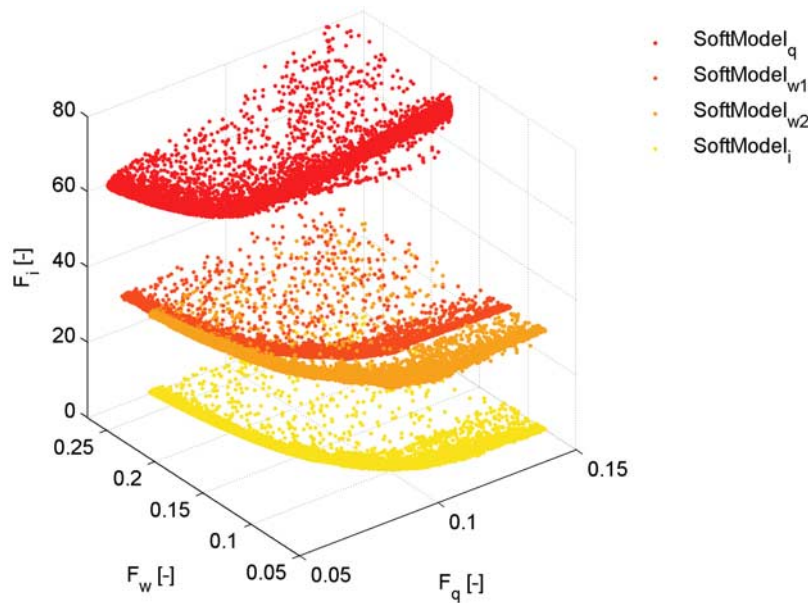
Figure 8. Performance of $SoftModel_q$, $SoftModel_{w1}$, $SoftModel_{w2}$ and $SoftModel_i$ with respect to discharge (F_q), groundwater level (F_w) and isotope concentration. Note that the improvements $SoftModel_{w1} \rightarrow SoftModel_{w2} \rightarrow SoftModel_i$ determine orthogonal movements in the objective space.

the reservoir levels can fluctuate below the threshold that produces runoff. Moreover, the lower reservoir is characterized by a nonlinear relation between storage and water level, which is motivated by an observed decrease of drainable porosity with depth. $SoftModel_i$, which describes the catchment behavior with respect to discharge, ground-

water, and isotope data, is characterized by an additional storage which accounts for the mixing of rainfall with water existing in the catchment prior to the event. Such storage is mostly identifiable with that of the unsaturated zone.

[51] The iterative process of model development allows associating model parameters with specific calibration
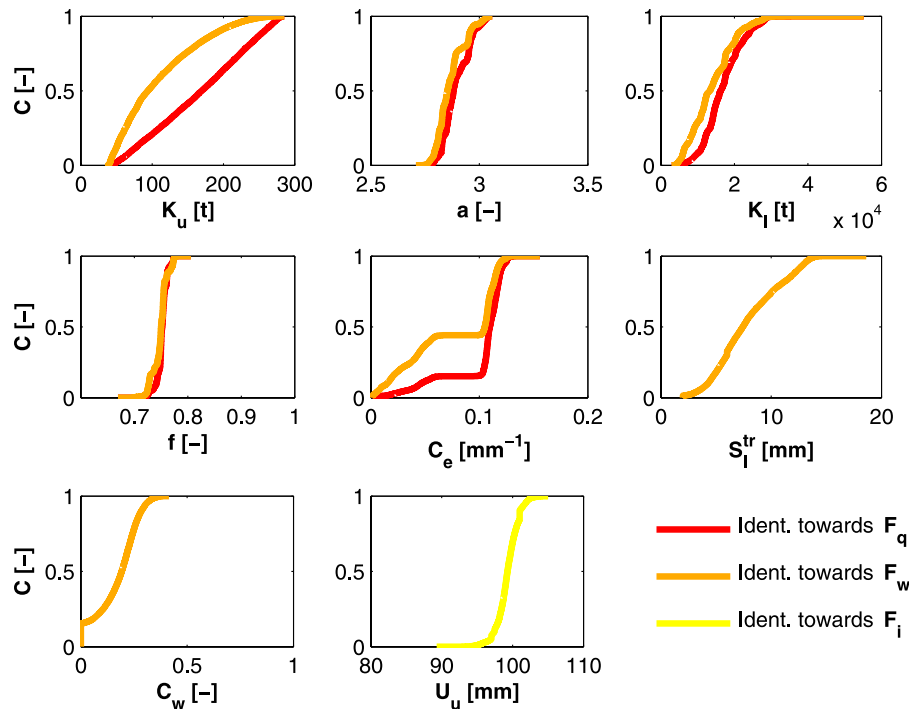

Figure 9. Identifiability of model parameters toward the related objectives. The timestep t is 20 min. The performance measures based on which the cumulative performance C is calculated are determined from the values of F_q, F_w, and F_i using the respective acceptance thresholds of 0.15, 0.3 and the maximum of F_i.

objectives. This idea is exploited in the calibration process, which is performed stepwise. It narrows the number of parameters for which, at each step, calibration is performed, and it improves the conditions for an efficient parameter sampling. Moreover, the fact that model parameters are associated to specific roles in the reproduction of the system response, allows a better control on model behavior. The sensitivity analysis shows that model parameters are generally well identifiable within relatively narrow ranges. This is a consequence of the fact that additional hypotheses about the catchment behavior are introduced only in correspondence of model failure. Hence the number of model parameters grows only as long as there are aspects of the system response that are not accurately represented, reducing the risk of model overparameterization.

5.2. Value of Different Forms of Data

[52] Our adaptive approach to model structure facilitates the dialog between modeler and experimentalist, and sheds light on the value of different forms of data in the model evaluation process. The developed models contain a minimal set of hypotheses to represent the available data. Most importantly, they stand a high chance of being rejected when evaluated with additional information [*Kirchner*, 2006]. This way, the value of additional information in illuminating new aspects of the catchment behavior becomes clear. It also demonstrates the value of "orthogonal" data; new, con-redundant measures of catchment behavior that the model is challenged to represent. Our work suggests that (1) discharge data provides information on the dynamics of storage (represented by the "free" water in the reservoirs) subject to pressure wave propagation generated by rainfall into the catchment, (2) groundwater data provides information on thresholds and on the contribution of different portions of the catchment to stream discharge, and (3) isotope data provides information on particle transport and mixing of the rainfall with the storage present in the catchment. Moreover, while groundwater data appear to be correlated with discharge data, and only a marginal improvement could be obtained adding this information to the model development process, isotope data appear to provide an orthogonal view on catchment behavior.

[53] Although this work does not provide new information on the behavior of the Maimai catchment, it arguably contributes to new understanding on the value of data for modeling. We use the Maimai catchment as a "proof-of-concept" location to investigate how different forms of data map to model components. While most catchments do not have the abundance of data that is available at Maimai, in the process of gauging an ungauged catchment it is important to have an indication on the potential contribution of data to process understanding. As *Bonell et al.* [2006] states: "We need data-rich basins to improve our ability to make reliable predictions in data-scarce basins". This work gives new guidance to what to measure, in what order and why, which are questions that are particularly important for the research in ungauged basins.

6. Conclusions

[54] We have presented an approach to the development an a la carte catchment rainfall-runoff model structure. The model was evolved iteratively as different types of data

were introduced and used to test its performance. The data used included measurements of stream discharge, groundwater levels and stream isotope concentration. The developed model was able to accurately reproduce the observed system response, and its parameters were identifiable given the calibration record and the selected measures of performance. We attributed this result to the iterative approach used. Unlike traditional one-size-fits-all model structures, we make the case for evolving model complexity in correspondence with observed model shortcomings. Our approach addresses the uniqueness of place issue in that the model reflects hypotheses that are related to the specific area, whereas a predefined model structure might be developed for different environmental conditions. This is a result of an interaction between modeler and experimentalist which is an integral part of this approach.

[55] Data were introduced progressively to allow evolving model complexity as they inform additional model components. A suitable structure for representing discharge dynamics was represented by a series of two reservoirs: an upper hillslope linear reservoir, and a lower hillslope nonlinear reservoir. The model evaluation with respect to discharge data required additional improvements, which necessitated the introduction of thresholds (characterizing the selective contribution of different portions of the catchment) and a nonlinear distribution of storage within the soil profile. The evaluation of this model with respect to isotope data required the introduction of additional storage which can be identified with the unsaturated zone. The ability of the evolving model to reproduce additional data allowed estimation of the contribution of complementary information to process understanding. While groundwater data appeared to be relatively correlated to discharge, stream isotope data appear to be independent, providing an orthogonal view on the catchment behavior.

References

Barnes, C. J., and M. Bonell (1996), Application of unit hydrograph techniques to solute transport in catchments, *Hydrol. Process.*, *10*(6), 793–802.

Bergström, S. (1995), The HBV model, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 443–476, Water Resour. Publ., Colo, ISBN 0-918334-91-8.

Beven, K. J. (2000a), *Rainfall-Runoff Modelling – The Primer*, John Wiley, Hoboken, N. J.

Beven, K. J. (2000b), Uniqueness of place and the representation of hydrological processes, *Hydrol. Earth Syst. Sci.*, *4*(2), 203–213.

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

Beven, K. J., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*, 279–298.

Birge, B. (2003), PSOt: A particle swarm optimization toolbox for use with MATLAB, in *Proc. IEEE Swarm Intelligence Symp. 2003*, pp. 182–186, Indianapolis, Indiana.

Bonell, M., J. J. McDonnell, F. N. Scatena, J. Seibert, S. Uhlenbrook, and H. A. J. van Lanen (2006), HELPing FRIENDs in PUBs: Charting a course for synergies within international water research programmes in gauged and ungauged basins, *Hydrol. Process.*, *20*, 1867–1874, doi:10.1002/hyp.6196.

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, vol. xviii, 588 pp., Addison-Wesley, Boston, Mass.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002), A fast elitist nondominated sorting genetic algorithm for multi-objective optimization: NSGA-II, *IEEE Trans. Evolutionary Comput.*, 6(2), 182–197.

Duan, Q. (2003), *Calibration of Watershed Models*, vol. vi, 345 pp., AGU, Washington, D. C.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031.

Engelbrecht, A. P. (2005), *Fundamentals of Computational Swarm Intelligence*, John Wiley, Hoboken, N. J.

Eriksson, E. (1971), Compartment models and reservoir theory, *Ann. Rev. Ecol. Syst.*, 2, 67–84.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2006), Is the groundwater reservoir linear? Learning from data in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 10, 139–150.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resour. Res.*, 43, W03434, doi:10.1029/2006WR005098.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, 44, W01402, doi:10.1029/2006WR005563.

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, 2161–2173.

Freer, J., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining Dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *J. Hydrol.*, 291, 254–277.

Grayson, R. B., and G. Blöschl (2000), Summary of pattern comparison and concluding remarks, in *Spatial Patterns in Catchment Hydrology: Observations and Modeling*, edited by R. B. Grayson and G. Blöschl, pp. 355–367, Cambridge Univ. Press, New York.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751–763.

Gupta, H. V., K. J. Beven, and T. Wagener (2005), Model calibration and uncertainty estimation, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson et al., pp. 2015–2032, John Wiley, Hoboken, N. J.

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Process.*, doi:10.1002/hyp.6989, in press.

Hornberger, G., and R. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, 12, 7–18.

Jothityangkoon, C., M. Sivapalan, and D. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, *J. Hydrol.*, 254, 174–198.

Kendall, C., and E. A. Caldwell (1998), Fundamentals of isotope geochemistry, chap. 2, in *Isotope Tracers in Catchment Hydrology*, edited by C. Kendall, and J. J. McDonnell, pp. 51–86, Elsevier, New York.

Kirchner, J. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.

Klemes, V. (1983), Conceptualization and scale in hydrology, *J. Hydrol.*, 65, 1–23.

Langbein, W. B. (1979), Overview of conference on hydrologic data, networks, *Water Resour. Res.*, 15, 1867–1871.

McDonnell, J., M. Bonell, M. K. Stewart, and A. J. Pearce (1990), Deuterium variations in storm rainfall: Implications for stream hydrograph separation, *Water Resour. Res.*, 26(3), 455–458.

McDonnell, J. J., M. K. Stewart, and I. F. Owens (1991), Effect of catchment-scale subsurface mixing on stream isotopic response, *Water Resour. Res.*, 27, 3065–3073.

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, New Zealand, *J. Hydrol.*, 257, 1–26.

Mosley, M. P. (1979), Streamflow generation in a forested watershed, *Water Resour. Res.*, 15, 795–806.

Moss, M. E. (1979a), Some basic considerations in the design of hydrologic data networks, *Water Resour. Res.*, 15, 1673–1676.

Moss, M. E. (1979b), Space-time and the third dimension (model. error), *Water Resour. Res.*, 15, 1797–1800.

Nash, J. (1960), A unit hydrograph study with particular reference to British catchments, *Proc. Inst. Civ. Eng., Parts 1 and 2*, 17, 249–282.

O'Kane, J. P. (2006), The hysteretic linear reservoir - A new Preisach model, *Physica B*, 372, 388–392.

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

Refsgaard, J. C., and B. Storm (1996), Construction, calibration and validation of hydrological models, in *Distributed Hydrological Modelling*, edited by M. B. Abbott and J. C. Refsgaard, pp. 41–54, Springer, New York.

Savenije, H. H. G. (2004), The importance of interception and why we should delete the term evapotranspiration from our vocabulary, *Hydrol. Process.*, 18(8), 1507–1511.

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.

Seibert, J., A. Rodhe, and K. Bishop (2003), Simulating interactions between saturated and unsaturated storage in a conceptual runoff model, *Hydrol. Process.*, 17, 379–390.

Sivapalan, M., and P. C. Young (2005), Downward approach to hydrological model development, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson et al., pp. 2081–2100, John Wiley, Hoboken, N. J.

Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Process.*, 17, 2101–2111.

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

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

Stewart, M. K., and J. J. McDonnell (1991), Modeling base flow soil residence times from deuterium concentrations, *Water Resour. Res.*, 27, 2681–2693.

Tang, Y., P. Reed, and T. Wagener (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, 11, 793–817.

Tromp-van Meerveld, H. J., and J. J. McDonnell (2006), Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, *Water Resour. Res.*, 42, W02411, doi:10.1029/2004WR003800.

Vaché, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resour. Res.*, 42, W02409, doi:10.1029/2005WR004247.

Wagener, T. (2003), Evaluation of catchment models, *Hydrol. Process.*, 17, 3375–3378.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stoch. Environ. Res. Risk Assess.*, 19, 378–387, doi:10.1007/s00477-005-0006-5.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26.

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. Process.*, 17, 455–476.

Weiler, M., and J. McDonnell (2004), Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology, *J. Hydrol.*, 285(1–4), 3–18.

Weiler, M., and J. J. McDonnell (2007), Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes, *Water Resour. Res.*, 43, W03403, doi:10.1029/2006WR004867.

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

Young, P. C., and M. J. Lees (1993), The active mixing volume: A new concept in modelling environmental systems, in *Statistics for the Environment*, edited by V. Barnett and K. F. Turkman, pp. 3–43, John Wiley, Hoboken, N. J.

————————————

F. Fenicia, Public Research Center – Gabriel Lippmann, 41 rue du Brill, L-4422, Belvaux, Luxembourg. (fenicia@lippmann.lu)

J. J. McDonnell and H. H. G. Savenije, Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, Netherlands.