Understanding catchment behavior through stepwise model concept improvement

Fabrizio Fenicia,[1,2] Hubert H. G. Savenije,[2] Patrick Matgen,[1] and Laurent Pfister[1]

[1] Lack of data is one of the main limitations for hydrological modeling. However, it is often used as a justification for over simplifying, poorly performing models. If we want to enhance our understanding of hydrological systems, it is important to fully exploit the information contained in the available data, and to learn from model deficiencies. In this paper, we propose a methodology where we systematically update the model structure, progressively incorporating new hypotheses of catchment behavior. We apply this methodology to the Alzette river basin in Luxembourg, showing how stepwise model improvement helps to identify the behavior of this catchment. We show that the most significant improvement of the evolving model structure is associated to the characterization of antecedent wetness. This is improved accounting for interception, which affects vertical storage distribution, and accounting for rainfall spatial heterogeneity, which influences storage variations in the horizontal dimension. Overall, our results suggested that, due to the damping effect of the basin, the description of fast catchment response benefits more from spatially distributed information than that of slow catchment response.

Citation: 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.

1. Introduction

[2] The main challenge of hydrological modeling is to express the response of a catchment in terms of its state variables and characteristics. This calls for a solution of the "closure" problem, which requires specifying the conditions that allow the closure of the undetermined system of balance equations [*Reggiani et al.*, 2000]. A major obstacle to this challenge is the lack of appropriate measuring techniques, which hampers the identification of the mechanisms underlying the rainfall-runoff transformation [*Beven*, 2006]. To circumvent this problem, internal catchment behavior has to be inferred by other means.

[3] One approach is to build a model of the catchment and to identify it with the natural system. This approach considers the model as a "virtual laboratory" [*Weiler and McDonnell*, 2004] which is used to perform experiments in a controlled environment. This approach has been widely used in hydrology to illuminate aspects of catchment behavior that are difficult to measure and that are poorly understood, such as the sensitivity of catchment response to soil properties [e.g., *Herbst et al.*, 2006], land cover change [e.g., *Eckhardt et al.*, 2003], and rainfall input [e.g., *Schuurmans and Bierkens*, 2007; *Arnaud et al.*, 2002]. While the advantage of this approach is that everything about the model is known and controllable, what limits its reliability is the unconditioned identification of the model

with the natural system, which implicitly disregards the role of conceptual model errors.

[4] An alternative approach is to infer internal catchment behavior using real data. Clearly, the use of real data should place more confidence in the outcome of a study. However, more often than not, the only measured variable available to test model assumptions is catchment discharge. Several authors indicated that in this case the essential features of catchment behavior can be reliably estimated with only a handful of parameters [*Beven*, 1989; *Jakeman and Hornberger*, 1993; *Hsu et al.*, 1995; *Young and Parkinson*, 2002]. Under such a complexity constraint it is difficult to explore the internal catchment behavior beyond very simplistic descriptions, and support models that retain enough empirical realism to be useful in addressing complex environmental problems [*Vaché and McDonnell*, 2006]. Indeed, most studies comparing complex versus simpler models [*Perrin et al.*, 2001; *Carpenter et al.*, 2001; *Refsgaard and Knudsen*, 1996; *Michaud and Sorooshian*, 1994] including those related to the Distributed Model Inter-comparison Project [*Reed et al.*, 2004] could not detect significant differences in performance between different types of models.

[5] Most of these studies, however, evaluated model performance with respect to discharge data alone based on a single statistical summary. This implies a loss of information which affects both the level of complexity that a model can support without incurring in identifiability problems [*Wagener et al.*, 2003], and the assessment of relative merits of individual models in an inter-comparison study.

[6] In the absence of direct measurements of individual processes, we propose a methodology to better explore internal catchment behavior based on a combination between the "top-down" approach to model development, and

---
[1]Public Research Center – Gabriel Lippmann, Luxembourg.
[2]Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, GA Delft, Netherlands.
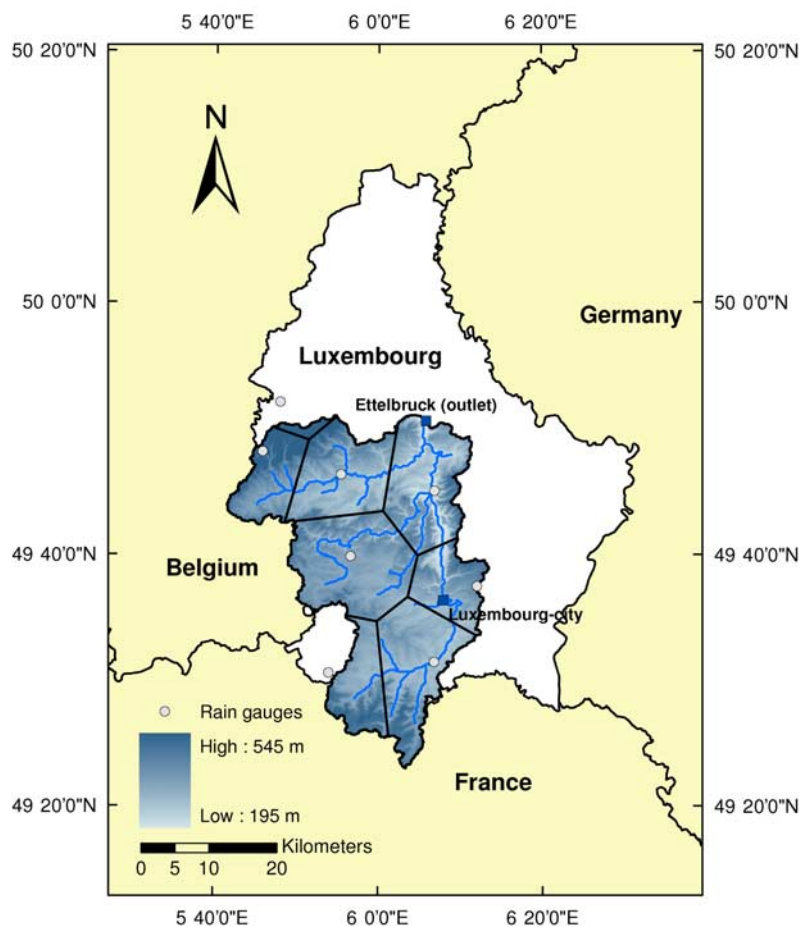
Figure 1. Location of the Alzette catchment in Luxembourg, raingauges location, and catchment subdivision in 8 Thiessen polygons.

a multiobjective approach to model evaluation. The top-down approach, initially introduced by *Klemes* [1983] and recently reformulated by *Sivapalan et al.* [2003], is based on a deductive philosophy which traces back the "causes" that are implicit in the overall "effect" produced by a system. In hydrological modeling, this approach would start from a simple structure which is progressively expanded based on its shortcomings in reproducing overall catchment behavior [e.g., *Jothityangkoon et al.*, 2001]. We use this approach in the process of "fingering down into the (smaller-scale) processes from above (i.e., catchment scale)" [*Sivapalan et al.*, 2003]. The multiobjective approach to model evaluation [*Gupta et al.*, 1998] is based on the consideration that a single measure of performance does not fully exploit the information contained in the data, and establishes that model performance should be evaluated with respect to several indicators. We use this approach not only to evaluate single models, but also to compare the performance of competing models [*Fenicia et al.*, 2007]. The combination of the two approaches provides a powerful tool to guide model improvement, and to evaluate different hypotheses on catchment behavior.

[7] We apply this methodology to the Alzette catchment in Luxembourg, where we test different model concepts first in a lumped and then in a distributed mode. In line with the top-down approach, we use successful model modifications to learn new aspects about the actual catchment behavior.

[8] One aspect of our results relates to the importance of the interception process. This process, while accounting for an important component of the water balance, is often neglected in model applications [*Savenije*, 2004]. With respect to discharge simulation, some authors found that the introduction of interception in a model structure improved model performance [e.g., *Zhang and Savenije*, 2005], others experienced deterioration of model performance [*Lindström et al.*, 1997].

[9] A second important aspect relates to the question of whether spatial heterogeneity of rainfall has an impact on catchment discharge, and for which combination of processes. While virtual experiments [e.g., *Arnaud et al.*, 2002] suggest that the catchment is sensitive to the spatial heterogeneity of rainfall, studies which employed real data [e.g., *Obled et al.*, 1994] support an opposite conclusion.

[10] The results of this work, as well as our considerations on the potential benefit of distributed information in catchment modeling, will be further analyzed in the discussion section of the paper.

2. Study Site and Data Description

[11] The study site is the Alzette river basin upstream of the Ettelbrück streamgauge. The catchment is mostly located in the Grand Duchy of Luxembourg and covers an area of about 1090 km^2 (Figure 1). The lithology is complex and
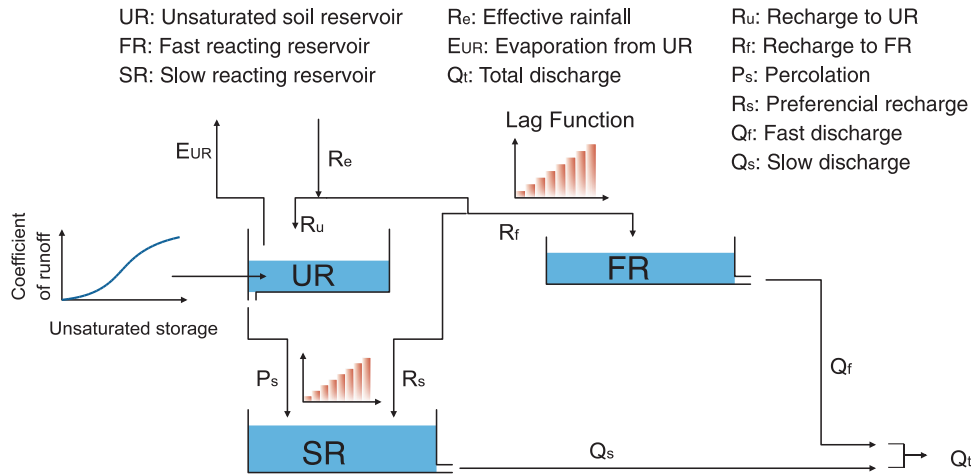
Figure 2. Schematic diagram of the $Flex_B$ model structure.

heterogeneous. The northern part is dominated by an impervious schist formation, the central part is characterized by marls, highly responsive to rainfall, and the southern part by sandstone and limestone, which constitute the main infiltration zone and the main groundwater reservoir of the basin. Land cover is composed of forest (34%), agriculture (23%), pasture (31%) and urban areas (12%).

[12] The catchment is instrumented with several recording raingauges. For this study, we selected 8 raingauges based on the available length of record (Figure 1). Rainfall and discharge data have been calculated at an hourly time step, while the daily potential evaporation (including all forms of evaporation) has been determined by the Hamon equation [*Hamon*, 1961] with data measured at Findel (Luxembourg airport). The daily potential evaporation has been disaggregated into hourly values using a sine curve distribution between the hours of sunlight. In total, five years of data were available for model evaluation, from September 2000 to August 2005. The last three years have been used for calibration and the first two for validation, as hydrograph peaks are somewhat larger in the calibration period.

[13] Within the catchment area, several plot scale experiments have been carried out to investigate small-scale catchment processes and their eventual link to larger scale catchment behavior. Of interest to this study are canopy interception measurements executed on a 0.06 ha experimental plot located in a central location within the basin (Huewelerbach catchment), at 380 m altitude. The plot is equipped with more than 80 pluviometers spaced at about a three-meter distance. Measurements were available for one hydrological year (2003–2004) and showed that the annual canopy interception rate is about 13% of the total incident precipitation. Beech litter interception was measured with a special forest floor interception device at ground level filled with forest litter [*Gerrits et al.*, 2007; *De Groen and Savenije*, 2006]. The box is equipped with weight sensors that allow continuous measurement of moisture storage. Preliminary results showed that litter interception (during the month of November) amounts to about 35% of the throughfall, indicating that the total annual evaporation from interception (by canopy and forest floor) is more than 43%.

[14] Interception data were not used directly for model evaluation, both because the forest floor interception data extend over a too short a period of time, and for the

difficulty of up-scaling the data to the entire basin. However, they were used as an indication of the order of magnitude of the interception process under the given climate and vegetation conditions.

3. Model Description

[15] We started modeling the catchment with a basic structure, based on the Flex model described by *Fenicia et al.* [2006]. The model is of the lumped conceptual type and is composed of several interconnected boxes representing key zones of catchment response. The basic version, which we named $Flex_B$, is characterized by three reservoirs: the unsaturated soil reservoir (UR), the fast reacting reservoir (FR) and the slow reacting reservoir (SR) (Figure 2). Rainfall input is first averaged with the Thiessen polygon method and then routed through the model. The model does not include a separate description of interception, hence rainfall (R) equals effective rainfall (R_e) which reaches UR. Infiltration into the soil (R_u) is evaluated based on a rainfall excess model that uses a distribution function of soil moisture storage capacity over the catchment (equations (1) and (2)).

$$C_r = \frac{1}{1 + \exp\left(\frac{-S_u/S_{fc}+1/2}{\beta}\right)} \tag{1}$$

$$R_u = (1 - C_r) \cdot R_e \tag{2}$$

[16] The coefficient of runoff C_r is a function of the storage in UR, S_u, and depends on the parameters S_{fc}, representing the maximum storage of UR, and a shape parameter β. When S_u is filled to capacity, excess rainfall is routed to FR. The part of rainfall that does not infiltrate either reaches SR through preferential recharge (R_s) or is routed as runoff (R_f) to FR (equations (3) and (4)).

$$R_s = (R_e - R_u) \cdot D \tag{3}$$

$$R_f = R_e - R_u - R_s \tag{4}$$

[17] Percolation is linearly related to the relative soil moisture content S_u/S_{fc} with a maximum value of P_{max} (equation (5)).

$$P_s = P_{max}(S_u/S_{fc}) \quad (5)$$

[18] As there is no interception, transpiration from vegetation and evaporation from interception are combined into a lumped evaporation term. This is often done in modeling, although conceptually it is considered an erroneous operation [*Savenije*, 2004]. In this approach, the actual total evaporation E_{UR} depends linearly on the relative soil moisture content until this ratio exceeds the threshold L_p, after which it equals the potential evaporation E_p (equation (6)).

$$E_{UR} = E_p \cdot \min\left(1, \frac{S_u}{S_{fc}} \frac{1}{L_p}\right) \quad (6)$$

[19] The fluxes entering FR and SR are lagged through triangular transfer functions of linearly increasing weights (Figure 2) that are defined by the parameters N_{lagf} and N_{lags}, which determine the number of time steps in the transformation routine. The fast and slow discharges Q_f and Q_s are calculated through a linear relation between storage and discharge (equations (7) and (8)).

$$Q_f = S_f/K_f \quad (7)$$

$$Q_s = S_s/K_s \quad (8)$$

[20] In total the model has nine parameters: a shape parameter for runoff generation β (−), the maximum UR storage S_{fc} (mm), the runoff partitioning coefficient D (−), the maximum percolation rate P_{max} (mm/h), the threshold for potential evaporation L_p (−), the lag times of the transfer functions N_{lagf} (h) and N_{lags} (h), and the timescales of FR and SR: K_f (h) and K_s (h).

4. Model Evaluation

[21] The issue of model evaluation has triggered a stimulating discussion in literature [*Wagener et al.*, 2003]. From the large variety of existing methods for model evaluation, we adopted the multiobjective optimization approach proposed by *Gupta et al.* [1998]. This approach evaluates a model simultaneously by a number of objective functions. The objective functions are defined so as to assess the model performance with respect to different aspects of the system's behavior, with lower values indicating better performance. The trade-off between the different performance indicators is visualized by the Pareto-optimal front, which identifies the maximum performance that can be achieved by the model, given the available data. Improvements in performance can therefore be easily tracked by a shift of the Pareto optimal front toward the origin of the objective function space [*Fenicia et al.*, 2007].

[22] The selection of performance measures can be based on different considerations, including knowledge of the errors involved in the modeling process, or, as in this case, arbitrary judgment related to the purpose of modeling [*Gupta et al.*, 1998]. The performance measures selected

for this study are the Nash and Sutcliffe coefficient C_{NS} and the correlation coefficient R. These functions increase with model performance, having a maximum value of one. Because optimization generally deals with a minimization procedure, the functions are subtracted from one, as shown in equations (9) and (10), yielding the objective functions F_{NS} and F_{CC}.

$$F_{NS} = 1 - C_{NS} = \frac{\sum_{i=1}^{n}(Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^{n}(Q_{o,i} - \overline{Q}_o)^2} \quad (9)$$

$$F_{CC} = 1 - R = 1 - \frac{\sum_{i=1}^{n}(Q_{s,i} - \overline{Q}_s)(Q_{o,i} - \overline{Q}_o)}{\sqrt{\sum_{i=1}^{n}(Q_{s,i} - \overline{Q}_s)^2 \sum_{i=1}^{n}(Q_{o,i} - \overline{Q}_o)^2}} \quad (10)$$

[23] Where n is the number of observations, i accounts for the time steps, and Q is discharge. The subscripts s and o indicate the simulated and observed values, while the overbar indicates an average value over the period of observation. F_{NS} has been chosen mostly for the purpose of communication, as it is a performance measure that is commonly used in modeling. This function is largely influenced by peak flow prediction errors due to the use of squared residuals. F_{CC} has been selected to complement F_{NS} with a relatively uncorrelated performance criterion. This function reaches a minimum value of zero when for all i: $(Q_{s,i} - \overline{Q}_s) = \alpha(Q_{o,i} - \overline{Q}_o)$, with α as a positive constant. It therefore indicates the agreement between the dynamics of the two hydrographs, disregarding the difference in their absolute values. In this case, it is useful to assess the time-shift between observed and simulated time series and serves the purpose of evaluating whether the timing is well captured by the model. The evaluation of this aspect is directly related to some of the modeling objectives (see section 5.2).

[24] As a sampling algorithm to estimate the Pareto-optimal front, the MOSCEM-UA (Multi Objective Shuffled Complex Evolution University of Arizona) algorithm [*Vrugt et al.*, 2003a] has been used. The algorithm requires the selection of a number of parameters: the maximum number of iterations (set at 20000), the number of complexes (set at 10), and the number of random samples that is used to initialize each complex (1000). Parameter bounds have been chosen based on the results obtained with a preliminary run of the MOSCEM-UA on a very large parameter space.

[25] Calibration results for the $Flex_B$ model structure are summarized in Figure 3. The figure shows the Pareto-optimal front associated to the $Flex_B$ model, and of other models that will be introduced further down. The Pareto-optimal front shows the trade-off between the selected modeling objectives. Each point on the front may not be considered better than any other point on the same front, since moving from one point to another corresponds to an improvement in one objective function and a deterioration in the other. Moreover, it is not possible to find points outside the front that assume better (lower) values for both objective functions. Therefore the Pareto-optimal front
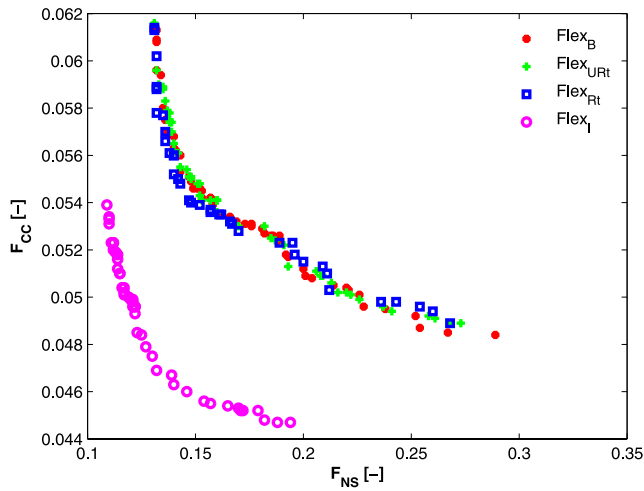
Figure 3. Model performance in the objective space. Pareto-optimal fronts of the Flex_B, Flex_{URt}, Flex_{Rt}, and Flex_I model structures.

marks the best performance that can be reached with a model with respect to the selected modeling objectives.

[26] The hydrograph generated by the Flex_B Pareto-optimal model with best F_{NS} value is represented in Figure 4. This model is characterized by a C_{NS} of 0.87, which is a relatively high value. Despite that the performance of the Flex_B model is relatively good, we investigated several modifications to this model structure to check whether they resulted in significant performance improvement. From this analysis, in line with the top-down approach, we tried to gain new insights into catchment behavior.

5. Proposed Modifications

5.1. Lumped Mode

[27] We tested a number of possible model improvements, first in a lumped, and subsequently in a distributed mode. In the lumped mode, we evaluated the effect of alternative model adjustments to account for threshold processes in runoff production. In the Flex_B model, the rainfall that produces fast runoff (routed to FR) depends merely on the value of C_r. Hence we tested the effect of three alternative hypotheses: (1) C_r is zero if UR is below a certain threshold UR_t (mm). Should this happen, all rainfall infiltrates into UR. This hypothesis was implemented in the Flex_{URt} model structure. (2) C_r is zero if rainfall (R) is below a threshold R_t (mm/h). This hypothesis was implemented in the Flex_{Rt} model structure. (3) We included an interception reservoir (IR) (Figure 5). The rainfall reaching IR fills the storage until the threshold $I_{\max}$ (mm) is reached. The amount that exceeds this threshold determines the effective rainfall R_e, which is subsequently routed through the same model as in Flex_B. During periods of no rain, water evaporates from IR at a rate E_{IR} equal to the potential evaporation E_p. E_{UR}, now to be regarded as transpiration from vegetation, is independent from the evaporation from interception E_{IR}, and is represented by equation (6), where L_p now represents the threshold for potential transpiration. We called this model Flex_I.

[28] All these model modifications involved the introduction of one extra parameter. The models were evaluated with the multiobjective approach described above. The results of the proposed model modifications are represented in Figure 3. If the model modification resulted in improved performance, the associated Pareto-optimal front would shift toward the origin of the axes. Despite that all proposed hypotheses are plausible, only the Flex_I model structure resulted in improved performance. The calibration routine set the optimal values of the parameters UR_t and R_t to zero. Hence the hypotheses represented by the Flex_{URt} and Flex_{Rt} model structures were not confirmed by the data, while the hypothesis contained in the Flex_I model structure helped to better explain catchment dynamics. In line with the top-down approach, this suggests that the interception process is in this case an important mechanism affecting catchment response. We note that both Flex_B and Flex_I can reproduce a correct water balance. In Flex_I, evaporation from interception E_{IR} was compensated by a reduction in transpiration E_{UR}, as the calibrated values of L_P were larger for Flex_I than
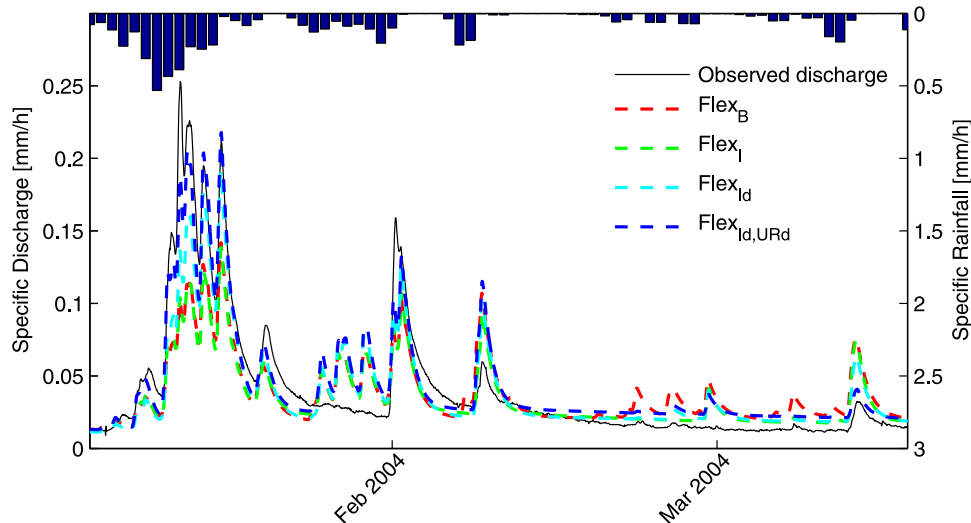

Figure 4. Hydrograph simulation (on a subset of the calibration record) of the Flex_B, Flex_I, Flex_{Id}, and $\text{Flex}_{Id,URd}$ Pareto-optimal models with best F_{NS}.
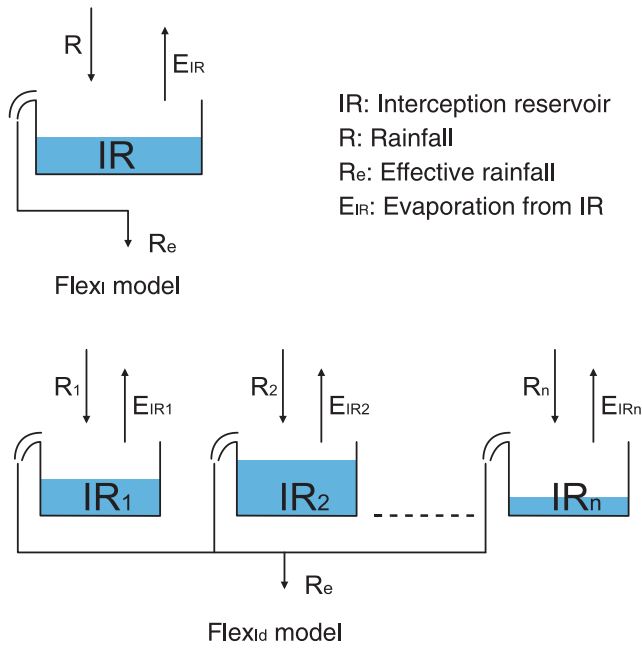
Figure 5. Schematic diagram of the interception component. The $Flex_I$ model implements a lumped interception reservoir, the $Flex_{Id}$ model accounts for distributed interception according to the raingauge location.

for $Flex_B$. Table 1 summarizes the contribution to the water balance of the output fluxes from $Flex_B$, $Flex_I$, and of other models that will be introduced further down. It can be observed that IR evaporates about 30% of the total rainfall volume.

[29] The hydrograph of the $Flex_I$ model with the best F_{NS} value is represented in Figure 4. It is possible to observe that while the simulation of peaks is similar to the one of $Flex_B$, the simulation of low flows is improved. The $Flex_B$ model produces secondary peaks during low flows that are not observed in reality, while the $Flex_I$ model is closer to observations. Apparently, the rainfall generating these small peaks in $Flex_B$ is in reality intercepted or added to the soil moisture (see Figure 4). While all proposed model modifications could in principle take this into account, the introduction of an interception store could act selectively on this aspect, without deteriorating model performance in other hydrograph regimes. In section 8.2 we will provide a more extended discussion on role of the interception process.

5.2. Distributed Mode

[30] While the previous model versions used average rainfall collected over the catchment, we subsequently tested whether accounting for the spatial heterogeneity of the rainfall could improve model results. We subdivided the catchment in n_T Thiessen polygons (Figure 1), to be regarded as model units, and subsequently stepwise introduced a distributed description of model components, starting with the interception reservoir, followed by the soil moisture reservoir and other system components. In doing so, we kept the rainfall for each Thiessen polygon as observed, without averaging. In this way the dynamics of the observed rainfall was maintained.

[31] The first modification consisted of considering for each of the n_T model units an interception reservoir (IR) of equal size. While in the other model versions rainfall was first averaged and subsequently routed through the model, in this case an individual, but equally sized, interception reservoir was assigned to each unit (Figure 5). The outputs of the different interception reservoirs are combined and weighted according to the area of each unit. Subsequently, the routing of the effective rainfall through the model proceeds as in the $Flex_I$ model, in a lumped manner. The different interception reservoirs have the same maximum threshold I_{max} and are characterized by the same evaporation rate E_{IR}. Therefore this model does not require more parameters than the $Flex_I$ model. However, at any time, each interception reservoir can have a different storage, depending on the associated rainfall history. We called this model structure $Flex_{Id}$.

[32] Subsequently, we distributed also the unsaturated reservoir (UR). Similarly, the distribution concerns only the reservoir internal states, and not the parameters. Output fluxes from UR are averaged and further processed in a lumped mode. We called this model structure $Flex_{Id,URd}$.

[33] While the previous structures did not use any information on the specific location of the individual raingauges, the subsequent model structures make use of this information. Through the digital elevation model of the study area, we calculated the average drainage distance to the outlet L (km) of each model unit, and we assumed a non linear relation between this distance and the lag time associated to the fast and slow reacting reservoirs:

$$N_{lagf,j} = T_f \left(\frac{L_j}{L_{max}} \right)^{\delta_f} \tag{11}$$

$$N_{lags,j} = T_s \left(\frac{L_j}{L_{max}} \right)^{\delta_s} \tag{12}$$

[34] Where j $(1, \ldots, n_T)$ indicates a model element, L_{max} (km) is the maximum of $L_{1, \ldots, nT}$, and T_f (h), δ_f (−), T_s (h), and δ_s (−) are the parameters characterizing N_{lagf} and N_{lags}, previously defined. The model structures that involve the distribution of lag times associated to FR, and to FR plus SR required one additional parameter each. These models were named $Flex_{Id,URd,LFd}$ and $Flex_{Id,URd,LFd,LSd}$ respectively.

[35] Distribution of the internal states of FR and SR was not considered because this would not bring any improvement. In fact, a set of linear reservoirs in parallel characterized by the same timescale would behave like a single linear reservoir. As we tried to maintain a low model

Table 1. Contribution to the Total Water Balance of the Output Fluxes of the $Flex_B$, $Flex_I$, $Flex_{Id}$, and $Flex_{Id,URd}$ Pareto-optimal Models With Best F_{NS}

	$Flex_B$	$Flex_I$	$Flex_{Id}$	$Flex_{Id,URd}$
E_{IR}	-	31%	29%	27%
E_{UR}	61%	35%	35%	37%
Q_f	17%	14%	15%	13%
Q_s	22%	20%	21%	23%

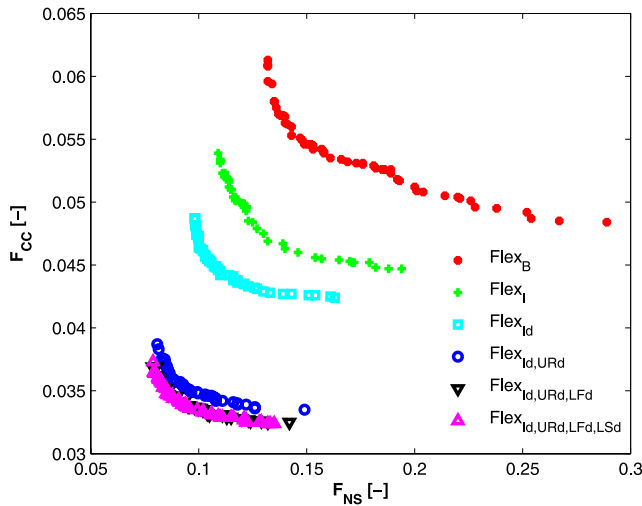**Figure 6.** Model performance in the objective space. Pareto-optimal fronts of the Flex$_B$, Flex$_I$, Flex$_{Id}$, Flex$_{Id,URd}$, Flex$_{Id,URd,LFd}$, and Flex$_{Id,URd,LFd,LSd}$ model structures.

complexity in terms of the parameter used, we did not attempt to distribute model parameters (besides internal states) for different catchment elements. While it is in principle possible to account for the heterogeneity of catchment characteristics in a conceptual and parsimonious way [*Refsgaard and Storm*, 1996], there is not much guidance on how to deal with this problem. Addressing this aspect would require an analysis that goes beyond the scope of this paper.

[36] The performance of the models in response to distributed rainfall input is represented in Figure 6. From the figure, it is possible to see that the Pareto-optimal front shifts progressively toward the origin of the axes, showing which model structural modifications help to better capture catchment dynamics. Flex$_{Id}$ performs better than Flex$_I$, indicating that interception, which is a threshold process, is particularly sensitive not just to rainfall volumes, but also to specific rainfall patterns. Flex$_{Id,URd}$ performs better than Flex$_{Id}$, which shows that the non-linear partitioning of effective rainfall into infiltration and fast runoff is also strongly affected by rainfall patterns. Flex$_{Id,URd,LFd}$ performs only slightly better than Flex$_{Id,URd}$. Apart from model structural errors, this may also indicate that the time to peak of the catchment does not depend much on the area where the rainfall is concentrated. This may be due to the shape of the catchment, which is not particularly elongated, or because in the study area the dominant direction of moving rainstorms is orientated perpendicularly to the direction of the main river, or because other factors, such as for instance lithology, may be more important than the distance of the rainfall event from the catchment outlet. Flex$_{Id,URd,LFd,LSd}$ performs like Flex$_{Id,URd,LFd}$, indicating that the distribution of lag times associated to slow flow does not result in any improvement. An explanation for this can be that slow processes, which are characterized by a large stock to flux ratio, are much less sensitive to rainfall spatial heterogeneity than fast processes. This point, as well as our considerations on the tradeoff between model complexity and model performance, will be further analyzed in the discussion section of the paper.

[37] Figure 4 shows the hydrographs of Flex$_{Id}$ and Flex$_{Id,URd}$ corresponding to the Pareto-optimal models with best F_{NS}. The hydrograph of subsequent structures are not reported as they do not introduce significant improvements, both with respect to objective functions, and in terms of visual inspection. It is possible to notice that while the introduction of an interception reservoir in lumped mode did not produce any improvement on hydrograph peaks, the introduction of distributed interception and the subsequent implementation of distributed soil moisture result in an improvement of peak runoff. This shows that the response of the catchment also during peaks is particularly sensitive to a correct description of the antecedent moisture condition, which depends on the spatial distribution of rainfall and interception.

6. Model Validation

[38] In order to check if the modifications introduced did not lead to an over-parameterization of the model structure, model performance was evaluated for a two years validation period (September 2000–August 2002). Model validation was performed by running the models Flex$_B$, Flex$_I$, Flex$_{Id}$, and Flex$_{Id,URd}$ with the parameter values of the respective Pareto-optimal fronts, using the time series of the validation period as input. Successive model structures were not considered, as they as they did not lead to significant improvements.

[39] The results are shown in Figure 7. We can see that the stepwise model improvement is confirmed by the validation, in the sense that the clusters of points corresponding to the different structures are progressively shifting toward the origin of the axes. This indicates that the structural modifications do not add unnecessary degrees of freedom to the model.

[40] Interestingly, the points computed for the validation period do not preserve the arc shape that is displayed during calibration (Figure 7). This is due to the presence of errors both in the model and in the data. Estimating these errors is a difficult task. However, the fact that an optimal model for the calibration period is not optimal during the validation period is a clear indication of their presence. In validation,
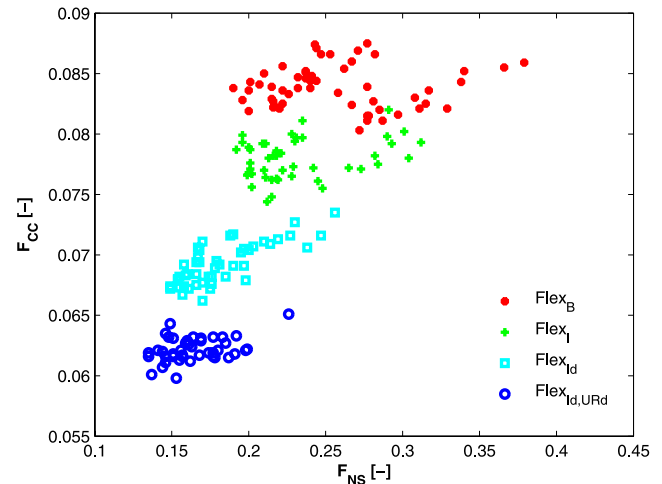

Figure 7. Performance of the Flex$_B$, Flex$_I$, Flex$_{Id}$, and Flex$_{Id,URd}$ models in validation mode.

Flex$_I$ did not perform better than Flex$_B$ in terms of C_{NS}. This may also be an effect of errors. The other models display better performance in both objective functions.

[41] While the effect of errors may put into question the role of optimization in hydrology, we still think that this approach is a powerful tool to evaluate model performance and compare different types of models. In the absence of information about the nature of the errors, it makes in fact sense to rely on data, and try to develop models that can closely reproduce them.

[42] With respect to the estimation of plausible parameter values, optimization may not be a good approach. Concentrating the selection of parameter values on the optimal models only may be too restrictive. Because of the presence of errors, in fact, a wider range of parameters may potentially be as good in representing the system. For this reason, the sensitivity analysis presented in the next section is not restricted to the Pareto-optimal models, but includes a larger number of potentially good models.

7. Sensitivity Analysis

[43] Sensitivity analysis deals with the question to what extent model outputs are affected by model parameters. Parameters with poor sensitivity cannot be well identified within the parameter space, and this is clearly a problem in hydrological modeling. If a model parameter cannot be identified, it means that there is little confidence in the model's correspondence with reality [*Kleissen et al.*, 1990]. A way to deal with this problem is to reduce model complexity to a level that can still be supported by the data [e.g., *Young and Parkinson*, 2002]. The main disadvantage of this approach is the risk of oversimplified conceptualization of the processes, which may not be useful for the purpose of modeling. An alternative approach is to find new (orthogonal) constraints [*Freer et al.*, 2004; *Vaché and McDonnell*, 2006], either through new measurements, or, as shown here, through an improved use of the available information.

[44] Many approaches exist that deal with the problem of estimating parameter sensitivity [*Tang et al.*, 2007]. In this paper, we adopted the approach described by *Freer et al.* [2004]. The approach is based on the RSA (Regional Sensitivity Analysis) of *Spear and Hornberger* [1980], which also forms the base of GLUE (Generalised Likelihood Uncertainty Estimation, *Beven and Binley*, 1992). *Freer et al.* [2004] applied this analysis in a multiobjective mode.

[45] Traditionally, the RSA is based on random sampling of the parameter space. This method however, is not efficient as it requires a large number of runs to cover all portions of the parameter space. Hence more efficient sampling strategies have been recommended [*Tang et al.*, 2007]. In this case, the sensitivity analysis was performed exploiting the same parameter sets generated by the MOSCEM-UA algorithm. The MOSCEM−UA algorithm integrates random sampling and Markov-Chain sampling, which allows a more efficient exploration of the parameter space in the region of the optima.

[46] The choice of objective functions can largely affect the outcomes of a sensitivity analysis. Most sensitivity analyses are based on the selection of a single statistical summary. This however, can underestimate the information content of the data, and consequently the influence of model parameters on model outcomes. Similar to the advantage of using multiobjective optimization, a multiobjective sensitivity analysis helps to better understand the role of model parameters on simulation. For consistence with the optimization procedure, we used the same (arbitrary) choice of objective functions selected for model optimization.

[47] The sensitivity analysis was performed on the Flex$_{Id,URd}$ model, which we consider the final outcome of the model development process. Subsequent model modifications were discarded, as they did not result in significant improvements. In order to evaluate the constraints that the two objective functions put on model parameters, the analysis was performed stepwise. In a first step, parameter sensitivity was evaluated with respect to F_{NS} only and was expressed determining the parameter samples that correspond to objective function values that are lower than the specified threshold $F_{NS,T}$. Subsequently, parameter samples were further constrained by discarding the parameter sets with values of F_{CC} that exceeded the threshold $F_{CC,T}$. $F_{NS,T}$ and $F_{CC,T}$ were arbitrarily specified at 0.15 (corresponding to a C_{NS} of 0.85) and 0.04 respectively. These values were selected in order to include the whole Pareto-optimal front and also sub-optimal models that are close to the Pareto-optimal frontier (Figure 6).

[48] The specification of a threshold value to separate between adequately and poorly performing models is at the heart of the RSA and of the GLUE approach [*Beven and Binley*, 1992]. This element of the analysis is based on the consideration that the position of optimal models in the parameter space is strongly dependent on errors in the data and in the model structure. Hence releasing the criterion of optimality in favor of a concept of "equifinality" [*Beven*, 1993] provides a better picture of the possible combinations that are equally well representative of catchment behavior. The results of the analysis are typically represented through scatterplots which are visually interpreted. Figure 8 displays parameter sensitivity with respect to F_{NS} only (grey dots), and to F_{NS} and F_{CC} simultaneously (black dots). Interestingly, the constraint on F_{CC} further reduces the sensitivity ranges for almost all parameters. The interception threshold I_{max} demonstrates poor sensitivity with respect to F_{NS} only, while it is further constrained when both objectives are used. The range of sensitivity of this parameter (from 1 to 4 mm) corresponds to an interception rate of up to 30% of the total rainfall, which is consistent with literature values [*Link et al.*, 2004] and observations in the catchment. The experiments performed in the study area showed a somewhat larger estimate of interception (see section 2). However, these measurements were executed on a land use type (beach forest) which is characterized by larger interception rates than the rest of the catchment. Moreover, due to the disparity of scales, the measurements are difficult to generalize to the whole catchment. N_{lags} shows no sensitivity at all with respect to both objective functions, showing that the processes related to this parameter may represent an unrealistic hypothesis, or that they cannot be assessed through the given combination of data and objective functions.

8. Discussion

8.1. On the "Art" of Modeling

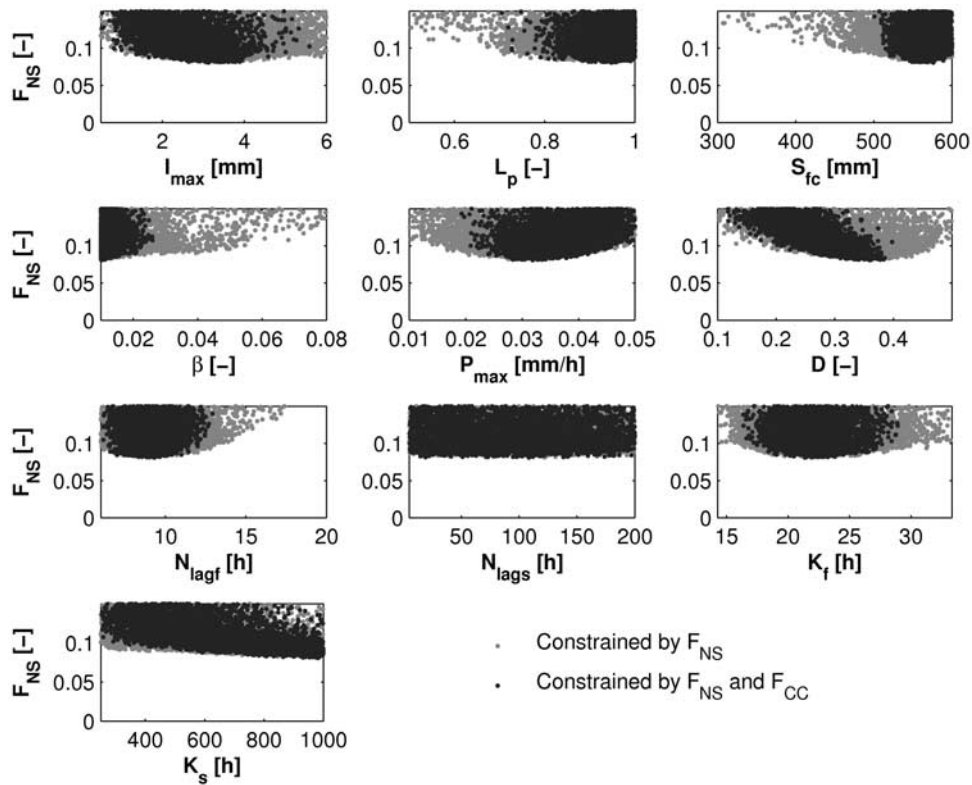[49] The main objective of hydrological modeling is to explain the variability of catchment response in terms of the

Figure 8. Scatterplots showing parameter sensitivity of the $Flex_{Id,URd}$ model structure to the selected objective functions (lower values of F_{NS} indicate better performance).

factors that may influence it. These factors include temporal and spatial variability of storage distribution, physical characteristics of the catchment, and forcing input. The relations that link the state variables to catchment response are often non-linear, hysteretic and scale-dependent, and their determination form one of the major problems that hydrological science is facing today [*Beven*, 2006].

[50] The effects of individual controls on catchment behavior have been mostly studied through virtual experiments [*Weiler and McDonnell*, 2004], which demonstrated that catchment response is sensitive to input variability, storage distribution, and catchment characteristics. However, in real cases, the evidence that catchment response can be easily parameterized using just a handful of parameters [*Beven*, 1989; *Jakeman and Hornberger*, 1993; *Hsu et al.*, 1995; *Young and Parkinson*, 2002] is already a demonstration that catchment behavior appears to follow simple laws [*Sivapalan et al.*, 2003; *Dooge*, 2005] depending of few controls.

[51] This contradiction is difficult to be solved with the current availability of observation techniques, which prevents us to shed light on the behavior of internal catchment processes. However, a possible way forward is to make better use of the data available, extracting in a more efficient way the information that it contains.

[52] In this study we identify a possibility to better explore internal catchment behavior in a combination between the top-down approach [*Sivapalan et al.*, 2003] and a multiobjective approach to model evaluation [*Gupta et al.*, 1998]. We show that the combination of these approaches forms a powerful tool to zoom-in from catchment to process

scale, and to test the effect of different hypotheses on catchment behavior.

[53] We demonstrate that increasing model complexity does not necessarily improve model performance (section 5.1). On the other hand, we also show that an improvement of model performance can be obtained just by improved use of the available information, without increasing the number of parameters (section 5.2). Hence it is not the number of model parameters that determines the model ability of reproducing catchment response. Instead, it is the role of these parameters, the processes they represent, and their impact on catchment response.

[54] This shows that modeling is both and Art and a Science. The science lies in the use of fundamental scientific principles and the formality of analysis; the art accounts for professional experience, insight, creativity and intuition. The latter is particularly important in developing a perceptual and conceptual model that captures the main processes at play, while maintaining minimum levels of complexity. As *Wagener et al.* [2003] points out "The modeler's task is to draw an inference from the type of failure that has occurred with respect to the hypothesis underlying the specific model component in order to develop an improved version". This inference is what we call the art of modeling.

8.2. On Interception as a "Dominant Process"

[55] Several studies underline the importance of interception as one of the key processes affecting the water balance. Vegetation intercepts a significant part of the rainfall, and in humid and temperate climates the amount of water that returns to the atmosphere as a result of canopy interception is estimated to vary between 10 and 50 percent of the annual

rainfall [*Link et al.*, 2004, and references within]. If also forest floor interception is taken into account, the amount of intercepted rainfall can be much larger [*Gerrits et al.*, 2007].

[56] Notwithstanding the important role of interception in the hydrological cycle, this process is often disregarded in hydrological models. Among them: the HBV model as described by *Bergström* [1995], the REW model [*Reggiani and Rientjes*, 2005], the Topkapi model [*Liu and Todini*, 2002], and the HYMOD model [*Vrugt et al.*, 2003b]. These models combine evaporation from interception and transpiration from vegetation into an "evapotranspiration" term. *Savenije* [2004] argues that this is a conceptual mistake, as interception has different dynamics from transpiration. The timescale of interception is short, as this process ends generally within a period of one day after rainfall [*De Groen and Savenije*, 2006], while transpiration has a much longer timescale due to the larger storage to flux ratio involved. The energy balance into the ground is in fact much smaller than on the surface, where due to air turbulence rapid rates of evaporation occur.

[57] Among the reasons for neglecting this process there certainly is the need to keep the model as simple as possible, so as to reduce the number of calibration parameters. Incorporating complexity into a model determines parameter identification problems and does not necessarily result in increased model performance [e.g., *Beven*, 2002]. Hence there is a need to focus on the "dominant processes" that mostly affect catchment response [*Grayson and Blöschl*, 2000].

[58] In some cases the implementation of interception has been related to water balance considerations [*Wagener et al.*, 2004]. However, hydrological models can easily produce a correct water balance compensating for structural errors by parameter readjustment [*Andréassian et al.*, 2004]. This is also shown by our work, where the introduction of evaporation from interception was compensated by a reduction of the transpiration flux. The effect of interception on the efficiency of discharge simulation is controversial. Some authors found that the introduction of an interception component in a model structure improved model performance [*Zhang and Savenije*, 2005], others experienced deterioration of model performance [*Lindström et al.*, 1997]. In both works, model performance has been assessed through C_{NS} only.

[59] In this study, we tested several threshold mechanisms for the production of runoff, and we demonstrated that the inclusion of the interception process improves model performance significantly. Hence in this case, interception, besides being conceptually important, is also a "dominant process" that affects catchment response. We note that in the model the interception component separates from the precipitation the effective part that contributes to soil moisture storage or runoff production. Hence while in the literature interception is often defined as a process whereby precipitation is intercepted by vegetation and subsequently evaporated before reaching the ground, in the present case, a more appropriate definition for this case is the one given by *Savenije* [2004] which considers interception as all the water that is intercepted by the wetted surface (including canopy, understorey, bare soil, litter, roads and build-up areas) and subsequently returned to the atmosphere during or shortly after a rainfall event.

[60] We link the improvement to a better description of antecedent wetness conditions. In the absence of an interception process, all rainfall is stored in the catchment or produces runoff. However, in reality, the infiltration and runoff production are threshold processes, which are highly sensitive to the wetness of soil surface and of the first centimeters of the soil. A more accurate representation of the storage in the soil and its spatial distribution, therefore, yields a better description of catchment response under variable wetness conditions.

[61] The divergence of opinions between this and other studies may be due to climatic conditions. In very wet or energy constrained conditions, a split between transpiration and interception may not improve model performance, while in drier conditions the effect of this process on the hydrograph will be more evident. Also, the importance of specific processes is strongly dependent on model evaluation criteria. In this case we built two performance measures to describe model agreement with observations (the Nash-Sutcliffe criterion and the correlation coefficient) and we showed that their combined use helps to better constrain parameter values. The use of additional orthogonal information to test model assumptions, such as, for instance, isotope concentration [*Vaché and McDonnell*, 2006], may shed more light on the relative importance of the interception process.

8.3. On the Effect of Rainfall Spatial Distribution on Catchment Discharge

[62] Spatial rainfall patterns can be highly heterogeneous. In general, short duration high intensity rainfall, typically generated by convective thunderstorms, has a very low spatial correlation, whereas the spatial correlation of long duration rainfall is larger. Whether or not spatial heterogeneity of rainfall has an impact on catchment discharge and for what reason, is a problem that has been often addressed in hydrology and that is still poorly understood. The question has been explored directly or indirectly in relation to several issues: the comparison of lumped versus distributed models, the eventual benefit of high density raingauge networks, the advantage of radar derived rainfall data at high spatial resolution, and the improved understanding of catchment internal processes.

[63] While the use of a model that mimics catchment behavior appears to be unavoidable, and is common to all studies addressing this question, the main difference lies in the use of fictitious versus real data. The majority of the works are "virtual experiments" [*Schuurmans and Bierkens*, 2007; *Arnaud et al.*, 2002; Koren et al., 1999; *Winchell et al.*, 1998; *Finnerty et al.*, 1997; *Krajewski et al.*, 1991; *Milly and Eagleson*, 1988; *Beven and Hornberger*, 1982; *Wilson et al.*, 1979] and their results seem to support the conclusion that catchment response is sensitive to spatial rainfall heterogeneity, depending on different processes and scales.

[64] There is much less work that addressed this question using real data, comparing model results with real observations. Interestingly, the majority of these studies seems to point into the opposite direction. While a correct estimate of rainfall volume has a significant effect on model predictions [*Andréassian et al.*, 2001; *Sun et al.*, 2000; *Obled et al.*, 1994] an accurate description of the rainfall spatial pattern does not always appear to be necessary to explain catchment

behavior. Most studies comparing distributed versus lumped models [*Carpenter et al.*, 2001; *Perrin et al.*, 2001; *Refsgaard and Knudsen*, 1996; *Michaud and Sorooshian*, 1994] including those related to the Distributed Model Inter-comparison Project [*Reed et al.*, 2004] could not detect significant differences in model performance. *Obled et al.* [1994] conclude that distributed rainfall estimates do not improve discharge predictions, hypothesizing that "the catchment has such a damping capacity that it does not require this type of information". *Smith et al.* [2004] support the hypothesis that only catchments with marked rainfall variability and little filtering effect will benefit from distributed rainfall information.

[65] Our results show that the spatial distribution of rainfall is relevant to explain catchment dynamics. Most importantly, the largest improvements in model performance are obtained without increasing the number of parameters. Hence the improvement is not an attributable of increased model complexity. The interception and unsaturated soil model compartments, which are characterized by a non-linear threshold-like behavior, are particularly sensitive to rainfall spatial distribution. Interestingly, information on the spatial distribution of individual raingauges (introduced by assuming a distribution of lag times associated to raingauge location) did not produce significant improvements. Apart from model structural errors, this may indicate that, in this catchment, spatial heterogeneity of rainfall does not have a strong impact on peak timing.

[66] In contrast to previous work analyzing this problem using real data, our outcome clearly shows that catchment response is sensitive to spatial heterogeneity of rainfall. One reason that may explain this difference could lie in the different set of hypotheses underlying alternative model structures. Another explanation could be that previous studies analyzed this problem on smaller catchment areas, where the spatial heterogeneity of rainfall is less evident. In general, the effect of rainfall spatial heterogeneity in modeling catchment behavior is clearly linked to model assumptions, catchment characteristics (e.g., catchment area and shape), climate, raingauge location and density. The effect of these controls needs further investigation in future research.

8.4. On the Utility of Distributed Information for Catchment Modeling

[67] While several authors indicated that the catchment acts as a low-pass filter attenuating the variability of the input signal [e.g., *Smith et al.*, 2004; *Andréassian et al.*, 2001; *Obled et al.*, 1994], we think that such damping effect is also related to the timescale of the processes affecting the rainfall-runoff transformation. Relatively fast processes, that is, those determining the fast response of the catchment to rainfall, demonstrate a higher sensitivity to the space-temporal variability of the input signal than relatively slow processes, such as those determining catchment response during dry weather periods. Our results show that substantial model improvement can be obtained through the distribution of interception, saturated soil, and to a minor extent the lag time of the fast reacting reservoir. All of these aspects contribute to the fast response of the catchment to rainfall. Interception is a process characterized by timescales in the order of several hours, and it is the first shackle in the chain of subsequent processes that determine catch-

ment response. The storage level in soil determines the proportion of effective rainfall which contributes to fast catchment response. No improvement at all is registered on the lag time distribution associated to slow processes.

[68] This result relates to the debate on the relative merits of distributed versus lumped models. As mentioned before, there is little agreement on whether distributed information helps to improve model predictions. In this respect, we think that the modeling of fast processes, like those related to surface or near-surface runoff, benefits more from distributed information than the modeling of slow processes, such as those related to groundwater flow. For fast processes, the variability of the forcing data is not reduced by the filtering effect of the catchment and is reflected in the variability of catchment response. For slow processes, the variability of the input variables is likely to be filtered out significantly by the averaging effect of the basin, so that long duration rainfall with a low spatial variability can be used, as is often the case, for groundwater related processes.

9. Conclusions

[69] In order to better understand internal catchment behavior, without the shortcomings of "virtual experiments", we adopted an approach that extracts information from real data in a more efficient way than is traditionally done. Our methodology is based on a combination of the "top-down" approach to model development, which is a framework for understanding catchment behavior based on data interpretation, and a "multiobjective" approach to model evaluation, which is based on the consideration that multiple measures of performance are needed to properly extract information from the data.

[70] The modeling started with a basic model structure applied to the Alzette catchment in Luxembourg. Subsequently, further refinements of model conceptualization were introduced and evaluated, initially in a lumped and then in a spatially distributed mode. We determined that model performance is particularly sensitive to the description of the state of wetness of the catchment. This may seem trivial, but we showed that the improved wetness strongly depends on the process of interception and on the distribution of model internal states in conjunction with distributed rainfall input. These results are of interest to ongoing discussions on which there is little consensus to date. In fact, the interception process, although accounting for an important component of the water balance, is often neglected in modeling application, particularly in relation to hydrograph simulation. Regarding the spatial heterogeneity of rainfall, while theoretical studies with artificial data show that it may have a considerable impact on catchment discharge, most applications using real data support an opposite conclusion.

[71] Our results contribute to the debate on the relative merits of lumped versus distributed models, showing that fast catchment response benefits more from distributed modeling than slow catchment response. Processes depending on a large stock to flux ratio (e.g., groundwater flow, transpiration, percolation) do not require information on the spatial distribution of rainfall, whereas fast processes, such as interception and surface runoff do. This is due to the damping effect of the basin, which filters the space-temporal

variability of the input signal and is larger for slow processes than for fast processes.

References

Andréassian, V., C. Perrin, C. Michel, I. Usart-Sanchez, and J. Lavabre (2001), Impact of imperfect rainfall knowledge on the efficiency of watershed models, *J. Hydrol.*, 250, 206–223.

Andréassian, V., C. Perrin, and C. Michel (2004), Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, *J. Hydrol.*, 286, 19–35.

Arnaud, P., C. Bouvier, L. Cisneros, and R. Dominquez (2002), Influence of rainfall spatial variability on flood prediction, *J. Hydrol.*, 260, 216–230.

Beven, K. (1989), Changing ideas in hydrology - The case of physically based models, *J. Hydrol.*, 10, 157–172.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51.

Beven, K. (2006), Searching for the Holy Grail of scientific hydrology: Qt = H (S, R, Δt)A as closure, *Hydrol. Earth Syst. Sci.*, 10, 609–618.

Beven, K. J. (2002), Uncertainty and the detection of structural change in models of environmental systems, in *Environmental Foresight and Models: A Manifesto*, edited by M. B. Beck, Elsevier Science, The Netherlands, 227–250.

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

Beven, K. J., and G. M. Hornberger (1982), Assessing the effect of spatial patterns of rainfall in modeling stream flow hydrographs, *Water Resour. Bull.*, 18(5), 823–829.

Carpenter, T. M., K. P. Georgakakos, and J. A. Sperfslage (2001), On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use, *J. Hydrol.*, 253, 169–193.

De Groen, M. M., and H. H. G. Savenije (2006), A monthly interception equation based on the statistical characteristics of daily rainfall, *Water Resour. Res.*, 42, W12417, doi:10.1029/2006WR005013.

Dooge, J. C. I. (2005), Bringing it all together, *Hydrol. Earth Syst. Sci.*, 9, 3–14.

Eckhardt, K., L. Breuer, and H. G. Frede (2003), Parameter uncertainty and the significance of simulated land use change effects, *J. Hydrol.*, 273(1–4), 164–176, doi:10.1016/S0022-1694(02)00395-5.

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, 139–151.

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.

Finnerty, B. D., M. B. Smith, D.-J. Seo, V. Koren, and G. E. Moglen (1997), Space–time sensitivity of the Sacramento model to radargage rainfall inputs, *J. Hydrol.*, 203, 21–38.

Freer, J. E., 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.

Gerrits, A. M. J., H. H. G. Savenije, L. Hoffmann, and L. Pfister (2007), New technique to measure forest floor interception – An application in a beech forest in Luxembourg, *Hydrol. Earth Syst. Sci.*, 311, 695–701.

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, Cambridge Univ. Press, Cambridge, 355–367.

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, 751–763.

Hamon, W. R. (1961), Estimating potential evapotranspiration: Proceedings of the American Soc. of Civil Eng., *J. Hydraulic Division*, vol. 87, no. HY3, p. 107–120.

Herbst, M., B. Diekkrüger, and J. Vanderborght (2006), Numerical experiments on the sensitivity of runoff generation to the spatial variation of soil hydraulic properties, *J. Hydrol.*, 326, 43–58.

Hsu, K., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, 31(10), 2517–2530.

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29(8), 2637–2649.

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.

Kleissen, F. M., M. B. Beck, and H. S. Wheater (1990), The identifiability of conceptual hydrochemical models, *Water. Resour. Res.*, 26(12), 2979–2992.

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

Krajewski, W. F., V. Lakshmi, K. P. Georgakakos, and S. C. Jain (1991), A Monte-Carlo study of rainfall sampling effect on a distributed catchment model, *Water Resour. Res.*, 27(1), 119–128.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, 201, 272–288.

Link, T. E., M. H. Unsworth, and D. Marks (2004), The dynamics of rainfall interception by a seasonal temperate rainforest, *Agric. Forest Meteorol.*, 124, 171–191.

Liu, Z., and E. Todini (2002), Towards a comprehensive physically-based rainfall-runoff model, *Hydrol. Earth Sys. Sci.*, 6(5), 859–881.

Michaud, J., and S. Sorooshian (1994), Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed, *Water Resour. Res.*, 30(3), 593–605.

Milly, P. C. D., and P. S. Eagleson (1988), Effect of storm scale on surface runoff volume, *Water Resour. Res.*, 24(4), 620–624.

Obled, C., J. Wedling, and K. Beven (1994), The sensitivity of hydrological models to spatial rainfall patterns: An evaluation using observed data, *J. Hydrol.*, 159, 305–333.

Perrin, C., C. Michel, and V. Andreassian (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchments model structures on 429 catchments, *J. Hydrol.*, 242(3–4), 275–301.

Reed, S. M., V. Koren, M. Smith, Z. Zhang, F. Moreda, and D.-J. Seo (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, 298(1–4), 27–60.

Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, 32(7), 2189–2202.

Refsgaard, J. C., and B. Storm (1996), Construction, calibration and validation of hydrological models, Distributed Hydrological Modelling, edited by M. B. Abbott and J. C. Refsgaard, Kluwer Academic Publishers, 41–54.

Reggiani, P., and T. H. M. Rientjes (2005), Flux parameterization in the representative elementary watershed approach: Application to a natural basin, *Water Resour. Res.*, 41, W04013, doi:10.1029/2004WR003693.

Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (2000), Conservation equations governing hillslope responses, *Water Resour. Res.*, 38(7), 1845–1863.

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

Schuurmans, J. M., and M. F. P. Bierkens (2007), Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, *Hydrol. Earth Syst. Sci.*, 11, 677–693.

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

Smith, M. B., V. I. Korena, Z. Zhang, S. M. Reed, J. J. Pan, and F. Moreda (2004), Runoff response to spatial variability in precipitation: an analysis of observed data, *J. Hydrol.*, 298(1–4), 267–286.

Spear, R. C., and G. M. Hornberger (1980), Eutrophication in Peel Inlet, II, Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, 14, 43–49.

Sun, X., R. G. Mein, T. D. Keenan, and J. F. Elliott (2000), Flood estimation using radar and raingauge data, *J. Hydrol.*, 239, 4–18.

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

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.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003a), Effective and efficient algorithm for multiobjective optimization

of hydrologic models, *Water Resour. Res.*, 39(8), 1214, doi:10.1029/2002WR001746.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003b), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters, *Water Resour. Res.*, 39(8), 1201, doi:10.1029/2002WR001642.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modeling: dynamic identifiability analysis, *Hydrol. Processes*, 17(2), 455–476.

Wagener, T., H. S. Wheater, and H. V. Gupta (2004), Rainfall-runoff modelling in gauged and ungauged catchments, Imperial College Press, London, UK, 330 pp.

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.

Wilson, C. B., J. B. Valdes, and I. Rodriguez-Iturbe (1979), On the influence of the spatial distribution of rainfall on storm runoff, *Water Resour. Res.*, 15(2), 321–328.

Winchell, M., V. H. Gupta, and S. Sorooshian (1998), On the simulation of infiltration and saturation excess runoff using radarbased rainfall estimates: effects of algorithm uncertainty and pixel aggregation, *Water Resour. Res.*, 34(10), 2655–2670.

Young, P. C., and S. Parkinson (2002), Simplicity out of complexity, in *Environmental Foresight and Models: A Manifesto*, edited by M. B. Beck, Elsevier Science, The Netherlands, 251–294.

Zhang, G. P., and H. H. G. Savenije (2005), Rainfall-runoff modelling in a catchment with a complex groundwater flow system: Application of the representative elementary watershed (REW) approach, *Hydrol. Earth Syst. Sci.*, 9, 243–259.

———————————————

F. Fenicia, P. Matgen, and L. Pfister, Public Research Center – Gabriel Lippmann, 41 rue du Brill, L-4422 Luxembourg.

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. (fenicia@lippmann.lu)