Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments

Ludovic Oudin,[1] Vazken Andréassian,[2] Charles Perrin,[2] Claude Michel,[2] and Nicolas Le Moine[2]

Received 8 June 2007; revised 20 November 2007; accepted 6 December 2007; published 14 March 2008.

[1] Given the contradictory results from recent studies, this paper compares classical regionalization schemes of catchment model parameters over the wide range of hydroclimates found in France. To ensure the generality of the conclusions, we used two lumped rainfall-runoff models applied to daily data over a large set of 913 French catchments. Three types of approaches were considered: regionalization using regression, regionalization based on spatial proximity and regionalization based on physical similarity. This comparison shows that in France, where a dense network of gauging stations is available, spatial proximity provides the best regionalization solution. The regression approach is the least satisfactory, with results very close to those obtained using one median parameter set for the whole country. The physical similarity approach is intermediary. However, the results obtained with these three methods lag far behind those obtained by full model calibration. Our results also show that some improvement could be made by combining spatial proximity and physical similarity, and that there is still considerable room for progress in the field of ungaged catchment modeling.

Citation: Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, *44*, W03413, doi:10.1029/2007WR006240.

1. Introduction

1.1. Regionalizing Rainfall-Runoff Models

[2] Predictive tools for water resources management are essentially data-driven, i.e., they need to be calibrated with observed flow data (the easiest case being when they are applied to gaged catchments). Major problems are encountered when applying these models to ungaged catchments for which no flow data are available and hence no such calibration is possible. Therefore especially since the 1970s, hydrologists have been attempting to develop strategies to estimate the parameters of their models without calibration [e.g., see *James*, 1972; *Magette et al.*, 1976]. The term regionalization has roots in the process of regime classification and catchment grouping [e.g., see *Pardé*, 1933; *Gottschalk et al.*, 1979] and was later extended in the rainfall-runoff modeling context to the transfer of parameters from neighboring gaged catchments (also called donor catchments) to an ungaged catchment. Nowadays, the concept of regionalization applies to all methods aimed at estimating model parameter values on any ungaged catchment in a definable region of consistent hydrological response. Three kinds of regionalization approaches are considered, each one with its specific advantages and inherent drawbacks.

[3] Probably the most popular is regionalization based on regression [e.g., see *Magette et al.*, 1976; *Young*, 2006]: it consists in developing a posteriori relationships between catchment descriptors (both physical and climatic) and model parameter values calibrated on gaged sites. Once these relationships have been established, one determines the parameters of an ungaged basin using its physical and climatic attributes. There are two hypotheses underlying this approach. First, it considers that a well-behaved relationship exist between the observable catchment characteristics and model parameters, whereas unfortunately, most models have been shown to have no unique set of parameters to define the best model fit to the flow response of a catchment, the values of these parameters being more or less dependent on the specific conditions of the calibration period and/or the possible input errors or inadequacies [e.g., see *Yapo et al.*, 1996; *Oudin et al.*, 2006a; *Perrin et al.*, 2007]. The second underlying hypothesis is that observable catchment descriptors chosen for regressions provide information relevant to the behavior of the ungaged catchment. Unfortunately, the spatial variability of the catchment characteristics and the problems in observing underground characteristics are a major obstacle in identifying hydrologically relevant catchment descriptors.

[4] The second kind of regionalization approach is based on spatial proximity and is among the earliest attempts to model ungaged catchments [e.g., see *Egbuniwe and Todd*, 1976; *Vandewiele et al.*, 1991]. It uses the parameter values

[1]UPMC Univ Paris 06, Sisyphe, Paris, France.
[2]Cemagref, Hydrosystems and Bioprocesses Research Unit, Parc de Tourvoie, Antony Cedex, France.

0043-1397/08/2007WR006240

calibrated for geographic neighbors, the justification for it being that the physical and climatic characteristics are relatively homogeneous within a region, so that neighbors should behave similarly.

[5] The third type of regionalization approach can be seen as a synthesis of the spatial proximity approach and the regression-based approach: regionalization is based on the similarity between an ungaged catchment and one or more gaged donor catchments [e.g., see *Burn and Boorman*, 1993; *McIntyre et al.*, 2005]. Basically, it transfers parameter information between catchments which are neighbors not geographically, but rather in terms of observable catchment descriptors. The catchment descriptors used are the same as those used for the regression-based approach. Hereafter, this third type of regionalization is referred as the physical similarity approach.

1.2. Why Compare Existing Regionalization Methods?

[6] In the past, many attempts have been made to determine which regionalization approach was the most appropriate, but the results were not clear since many of these studies were based on relatively small catchment data sets, e.g., the twelve catchments used by *Goswami and O'Connor* [2006], making the conclusions virtually site-specific. Recent studies on larger data sets still do not seem to reach a consensus (see the review by *Merz et al.* [2006]). Here are four typical examples of disparate results obtained on larger catchment sets:

[7] • *Merz and Blöschl* [2004] and *Parajka et al.* [2005] tried to regionalize an 11-parameter version of the HBV model. Their analysis, based on more than 300 Austrian catchments, showed that regionalization based on spatial proximity with transposition of an entire set of parameters offers the best model performance on ungaged catchments. They also note that using nested catchments as donors may significantly improve the basic spatial proximity approach.

[8] • *McIntyre et al.* [2005] proposed a methodology based on physical similarity in a multimodel framework, i.e., averaging outputs obtained with model parameters from similar gaged catchments. The model tested was a five-parameter version of the PDM model applied on 127 UK catchments. In this study, the physical similarity approach outperformed the regression approach. Unfortunately, the authors did not include a comparison with a spatial proximity approach.

[9] • *Young* [2006] tried to regionalize a six-parameter version of the PDM model on 260 UK catchments. He showed that regressions relating individual model parameters and catchment attributes yield the most satisfactory results, compared to other approaches, including some based on spatial proximity and physical similarity.

[10] • *Kay et al.* [2006] compared the performance of three alternative methods of regionalization, for two six-parameter conceptual models (PDM and TATE) over 119 catchments in the UK, specifically for use in flood frequency estimation using continuous simulation. They found that the physical similarity method performs slightly better than the regression method for PDM, whereas regression performs best for TATE. The authors concluded that the best choice of a regionalization method is model-dependent.

[11] The reasons why the above-mentioned studies produced such different results are varied. First, studies were conducted on different catchment sets, and even if they are quite large, the conclusion can still be influenced by the data set used. Secondly, as pointed out by *Merz et al.* [2006], the available catchment descriptors differ from one study to another: for example, for the UK catchments, a classification of soil types provided additional critical physical information on soil hydrology, contrary to the Austrian case studies. Thirdly, the methodologies used in each type of approach depend on a number of unavoidable arbitrary choices made by the authors, which can influence the performance of the regionalized model (number of donor catchments, density of gaged catchments, use of poorly modeled catchments as donors). Last, as demonstrated by *Kay et al.* [2006], the performance of each regionalization approach may depend on the structure and parameterization of the rainfall-runoff models used.

1.3. Scope of the Paper

[12] This paper is a contribution to the search for the most appropriate regionalization strategy. Its originality lies in:

[13] • The catchment set of large size (913 French catchments), in a widely varying hydroclimatic context, in order to determine the directions to take for further research on ungaged catchment parameterization.

[14] • The investigation of the impact of model parameterization on the performance of regionalization schemes, which hitherto has not been studied on large sets of catchments, apart from the study by *Kay et al.* [2006] based on 119 UK catchments (see section 2).

[15] • The extent and depth of the comparison: three types of approaches were considered (see section 3): regionalization with regression, regionalization based on spatial proximity and regionalization based on the catchment characteristics' similarity.

[16] In section 4, we determine the optimal settings of the three regionalization schemes and we compare their efficiency in predicting daily streamflows on ungaged sites. We also discuss possible ways to improve the tested approaches. Last, in section 5, the findings are summarized.

2. Material

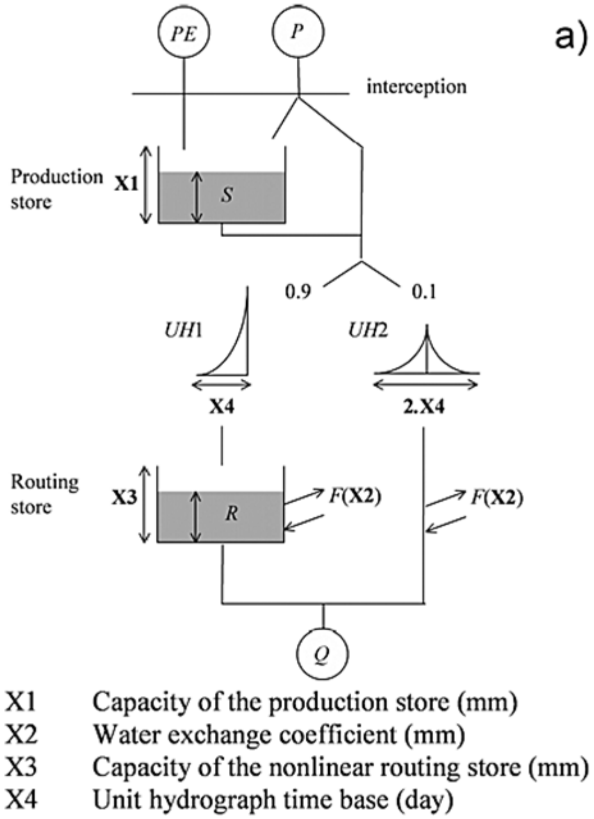
2.1. Rainfall-Runoff Models

[17] Two daily, continuous lumped rainfall-runoff models were used:

[18] • The GR4J rainfall-runoff model, an efficient and parsimonious (four free parameters) daily lumped continuous rainfall-runoff model. *Perrin et al.* [2003] provide a detailed description of the model.

[19] • The TOPMO rainfall-runoff model (six free parameters), inspired by TOPMODEL [*Beven and Kirkby*, 1979; *Michel et al.*, 2003], already tested on large data sets. This lumped model has the advantage of being quite different from GR4J but has comparable performance [e.g., *Oudin et al.*, 2006b]. Here the distribution of the topographic index is parameterized and optimized, and not calculated from a DEM. This was found to have only a limited impact on model efficiency, as shown by *Franchini et al.* [1996] and this eases the application of the model when it is tested on several basins.

[20] On gaged catchments, GR4J and TOPMO use four and six free parameters (see Figure 1) respectively. The optimization algorithm used to calibrate parameter values is a local gradient search procedure [*Edijatno et al.*, 1999].



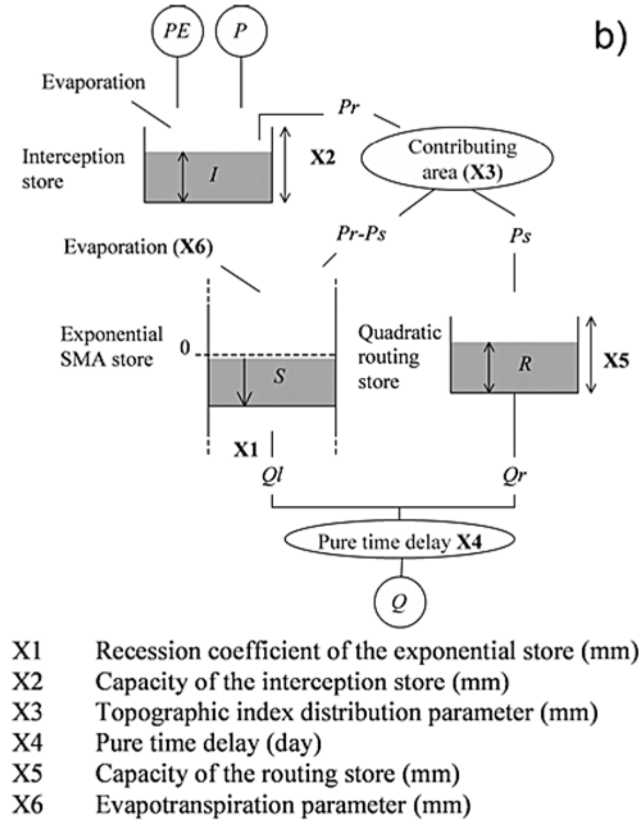


Figure 1. Schemes of (a) the GR4J and (b) the TOPMO rainfall-runoff models (PE: potential evapotranspiration; P: precipitation; Q: streamflow; Xi: ith model parameter; other letters are internal state variables).

The optimization starts from a default parameter set that is the most likely when the model is used on a large number of catchments. Then the optimization procedure searches step by step in the log-transformed parameter space in the direction that most improves the objective function. During the search, the search step progressively reduces to refine the location of the optimum. The search stops when the search step is below a given threshold, i.e., when one considers that the optimum was located with acceptable precision. This method was tested in several studies and showed good performances for models having up to eight parameters to calibrate [*Nascimento*, 1995; *Perrin*, 2000; *Mathevet*, 2005], including the models tested here. As the objective function, we used the *Nash and Sutcliffe* [1970] criterion computed on root square transformed flows ($\sqrt{Q}$), which was shown to be a good compromise between several alternative criteria [*Oudin et al.*, 2006b].

2.2. Catchment Set and Data Used

[21] We used a database of 913 French catchments located throughout France (see Figure 2) for which daily rainfall and runoff time series over the 1995–2005 period were available. Potential evapotranspiration time series were computed using air temperature data [see the equation developed by *Oudin et al.*, 2005]. Hydrometeorological conditions and physical catchment characteristics are quite diverse: the database contains 162 small Mediterranean catchments as well as larger, groundwater-dominated basins or highland catchments. Catchments where snow has a major hydrological role were excluded from our catchment set because we did not wish to use any snowmelt modules here so as to keep the models as parsimonious as possible. Five physical descriptors of the catchments were considered for regionalization purposes: catchment area (A_c), mean slope (S_c), median altitude (Z_c), river network density (d_c), and fraction of forest cover (f_c). We also included a hydroclimatic descriptor for the physical similarity approach: the aridity index (E_m/R_m), i.e., the ratio of mean annual potential evapotranspiration to mean annual rainfall. Other potentially relevant characteristics, such as soil types, were not available over this catchment set. Table 1 gives an overview of the distributions of the hydro-climatic characteristics of the catchment as well as the descriptors used for regionalization.

3. Tested Regionalization Schemes

3.1. Regionalization Using Regression

[22] This approach is the most popular among regionalization methodologies. To establish the regression equations, we only used those catchments for which model performance was judged acceptable, i.e., with a Nash-Sutcliffe criterion in calibration above 0.70, as we considered that poorly modeled catchments would not provide sufficient relevant information. In section 4.2, we will study the impact of this choice on the performance of the regionalized model. The coefficients of determination of the regression equations between the model parameter values and the six descriptors (A_c, S_c, Z_c, d_c, f_c, E_m/R_m) do not exceed 0.5. While extremely low, this is consistent with

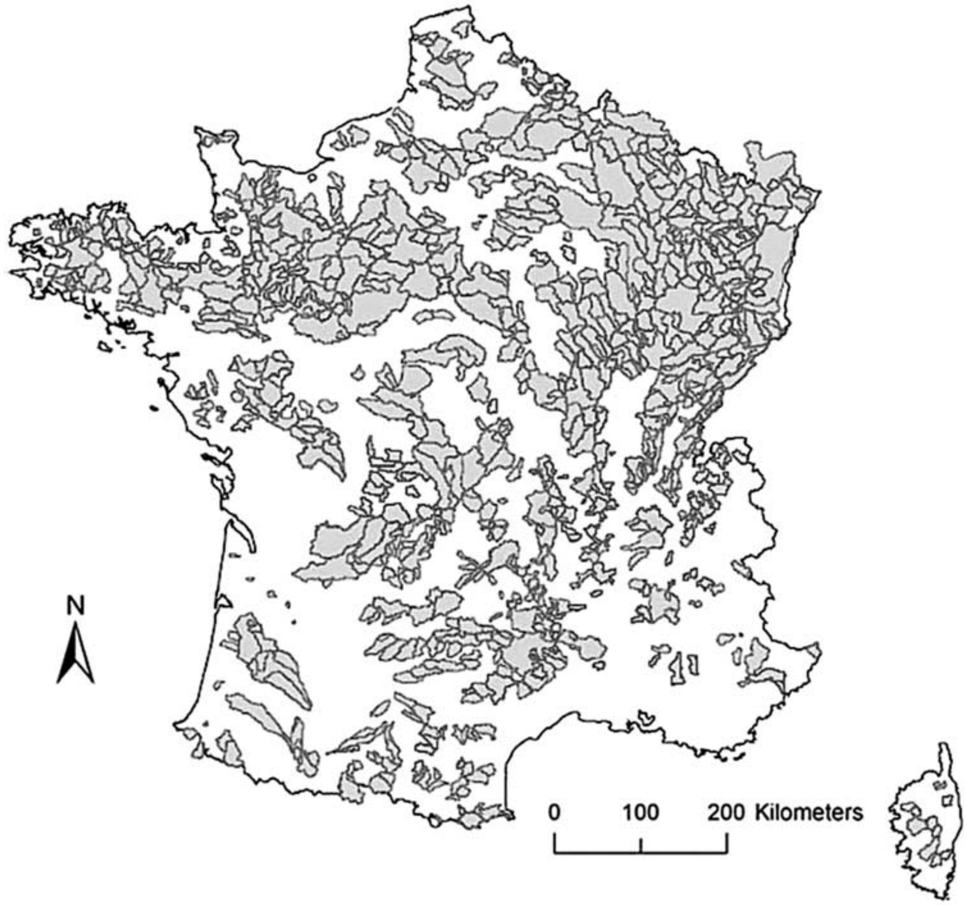


Figure 2. Location of the 913 basins used in this study.

the results of other regionalization studies on large catchment sets [*Merz et al.*, 2006]. Note that few of the above catchment descriptors are actually warranted by a significant Student t ratio, but all descriptors were used in the initial regression equations for the sake of uniformity.

[23] One well-known limitation of the regression approach is that model parameters may not be well defined (due to interactions between parameters): different parameter sets yielding virtually the same model performance may exist and therefore the regression equations obtained are likely to be weak. To reduce this possible bias, *Hundecha and Bárdossy* [2004] and *Bárdossy* [2007] proposed an alternative method by first assuming the functional form of the relationship between the catchment characteristics and the model parameters and then calibrating the model for many catchments simultaneously. This type of method was used for our catchment set with the GR4J model, but the results (not reported in this paper) showed only a very slight improvement compared to classical regression equations.

3.2. Regionalization Using the Spatial Proximity Approach

[24] The spatial proximity approach consists of transferring parameters from neighboring catchments to the ungaged catchment, the rationale being that catchments that are close to each other should have similar behavior since climate and catchment conditions should vary evenly in

Table 1. Distributions of Hydro-Climatic Characteristics and Descriptors of the 913 French Catchments Used

Catchment Characteristics	Notation and Unit	Min	10th Percentile	Median	90th Percentile	Max
Hydro-Climatic Characteristics						
Mean annual rainfall	*Rm* (mm/yr)	662	780	978	1488	2182
Mean annual potential evapotranspiration	*Em* (mm/yr)	339	587	659	738	845
Mean annual streamflow	*Qm* (mm/yr)	31	166	344	810	3493
Runoff yield	*Qm*/*Rm* (−)	0.03	0.21	0.34	0.58	4.24
Descriptors Used						
Catchment area	*Ac* (km^2)	10	37	148	918	9390
Catchment slope	*Sc* (−)	0.001	0.018	0.050	0.196	0.493
Median catchment altitude	*Zc* (m)	24	97	314	1024	2222
Drainage density	*dc* (−)	0.001	0.029	0.054	0.074	0.103
Fraction of forest cover	*fc* (−)	0.00	0.05	0.29	0.68	0.97
Aridity index	*Em*/*Rm* (−)	0.23	0.40	0.68	0.90	1.20

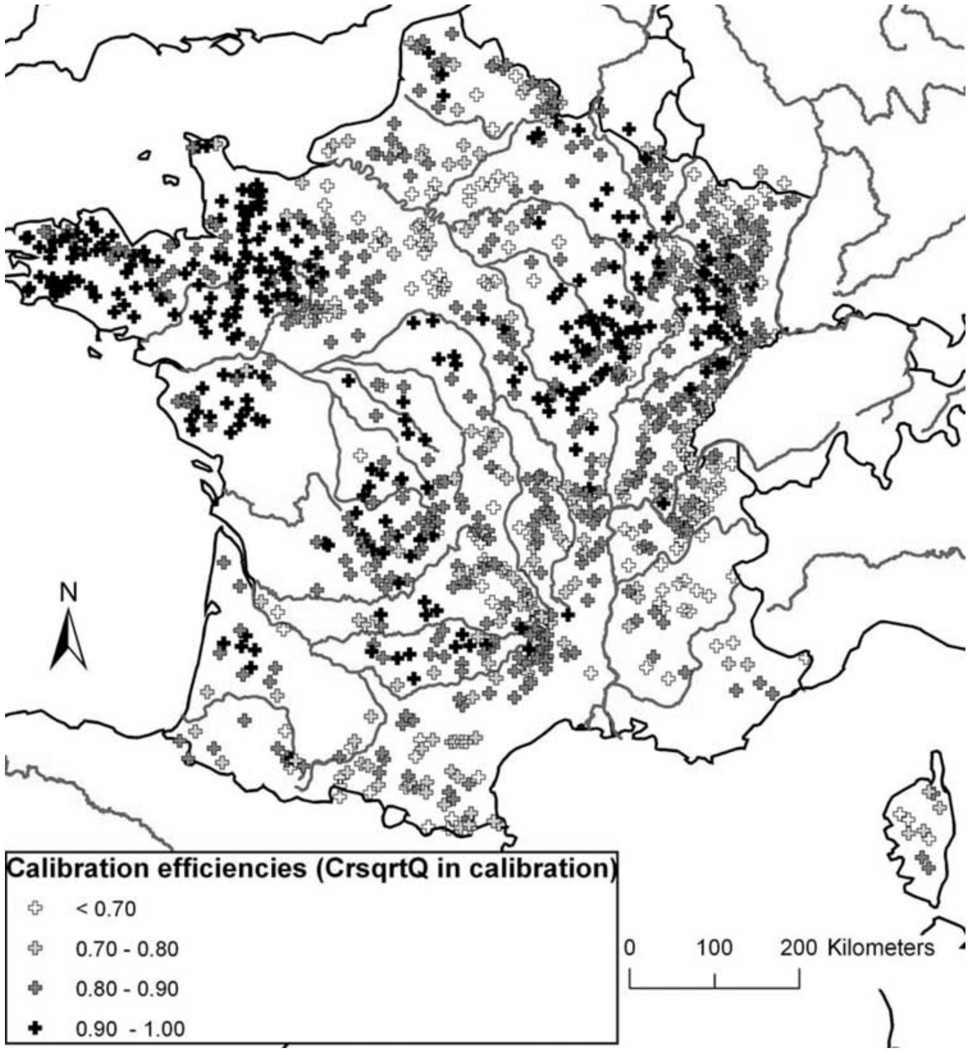


Figure 3. Calibration efficiencies over the 913 unregulated French catchments. GR4J model.

space. This approach is intuitively attractive, although it obviously depends on the density of the gaged basin network. Note that many refinements of this approach have been proposed, such as inverse distance weighting applied to the donor catchments, kriging of model parameter values [see, e.g., *Parajka et al.*, 2007], regional pooling for model calibration [*Goswami et al.*, 2007]. However, in this paper, we focus on the basics of this approach. The selection of the donor catchments is based on the proximity of the ungaged catchments to the gaged ones. We used the distances between catchment centroids.

[25] In this paper, we consider two options to combine the information from donor catchments:

3.2.1. Option 1: Parameter Averaging

[26] A regional set of parameters is computed as the mean of parameters from donor gaged catchments applied to the ungaged catchment. Thus streamflow for day j is computed as:

$$\hat{Q}(j) = \hat{Q}\left(j, \frac{\sum_{i=1,m} X_i}{m}\right) \quad (1)$$

where m is the number of donor (gaged) catchments considered and X_i is the vector of model parameter values for the donor catchment i.

3.2.2. Option 2: Output Averaging

[27] A regional computed streamflow is derived from the simulations obtained with the sets of parameters of the donor catchments. Thus streamflow for day j is computed as:

$$\hat{Q}(j) = \frac{1}{m} \sum_{i=1,m} \hat{Q}(j, X_i) \quad (2)$$

where m is the number of donor (gaged) catchments and X_i is the vector of parameters for the donor catchments i.

[28] The output averaging option, used by, e.g., *McIntyre et al.* [2005] differs from the parameter averaging option since it uses the unmodified parameter sets from the gaged catchments to the ungaged one, the advantage being that the full information content of the locally calibrated model parameters is used. Note however that when using only one gaged neighbor (m equal to unity), the two options are similar. These two options will be tested, as described in section 4, and the optimal number of donor catchments m will be determined.

Table 2. Performance of Rainfall-Runoff Models Over the Catchment Set[a]

Model	Assessment Criterion	Calibration		Verification	
		1996–2000	2001–2005	1996–2000	2001–2005
GR4J	Cr_Q	0.82	0.82	0.78	0.78
	$Cr_{\log Q}$	0.84	0.82	0.81	0.80
	Cr_{sqrtQ}	0.86	0.85	0.83	0.82
TOPMO	Cr_Q	0.78	0.78	0.71	0.73
	$Cr_{\log Q}$	0.79	0.79	0.73	0.74
	Cr_{sqrtQ}	0.82	0.82	0.76	0.77

[a]Median efficiency.

3.3. Regionalization Using the Physical Similarity Approach

[29] The physical similarity approach consists of transferring hydrological information from donor (gaged) catchments that are similar to the ungaged catchments in terms of catchment descriptors. The idea is not new and originates from basin classification studies [*Acreman and Sinclair*, 1986; *Nathan and McMahon*, 1990], the rationale being that catchments with similar attributes should behave similarly. The methodology used here is very close to the methodology developed and tested by *Rojas-Serna* [2004], who combined an approach based on physical similarity with calibration on point flow measurement. It is also very close to the approach followed by *McIntyre et al.* [2005], who also included prior likelihood based on the GLUE framework [*Beven and Freer*, 2001].

[30] Here, the selection of the donor catchments is based on the proximity of the ungaged catchments to the gaged ones in terms of catchment descriptors. Since catchment descriptors have different units and ranges, the donor catchments were ranked by decreasing proximity to the pseudo-ungaged catchment for each descriptor. When several descriptors were used, we used the ranks of the donor catchment for each descriptor to compute a mean rank that was then used to rank the donor catchments by decreasing proximity. Hence each descriptor had the same weight in the proximity computation.

[31] All possible combinations of one to six characteristics among the six descriptors mentioned in section 3.1 (A_c, S_c, Z_c, d_c, f_c and E_m/R_m) were tested. For a given ungaged catchment and a selected catchment descriptor, we looked for the most similar m catchments among the 912 gaged catchments. Then two options were considered: averaging the model parameter values of the similar m catchments (equation (1)) or averaging the outputs of the m gaged catchments (equation (2)).

3.4. Assessment of Regionalization Methods

[32] To assess the relative performance of the methods for flow estimation in ungaged catchments, each of the 913 catchments was used in turn as if it were ungaged, following a jack-knife procedure (also called pseudo-ungaged catchment procedure). After simulation of its flow by applying a regionalization method, its observed streamflow time series was used to assess the efficiency of the regionalization procedure. Except for the preliminary tests described in section 4.1, tests were made over the entire 1995–2005 period (the year 1995 being used for model warm-up).

Table 3. Correlation Coefficient (R) Between Parameters Calibrated on Two Different Subperiods (1995–2000 and 2000–2005)

	X1	X2	X3	X4	X5	X6
R for GR4J	0.75	0.80	0.86	0.78	–	–
R for TOPMO	0.40	0.36	0.47	0.72	0.29	0.73

[33] Three criteria were used to assess model efficiency on the verification periods. All of them are of the least squares type and based on the *Nash and Sutcliffe* [1970] criterion. One is the classical Nash-Sutcliffe efficiency (Cr_Q) and, of the other two, one is based on the same formulation but computed on square root transformed flows (Cr_{sqrtQ}) and the other on logarithmic transformed flows ($Cr_{\log Q}$). Cr_Q puts more emphasis on high-flow simulations, $Cr_{\log Q}$ on low-flow simulations, and Cr_{sqrtQ} gives an intermediate, more balanced picture of the overall hydrograph fit.

4. Results

4.1. Model Performance and Calibrated Parameter Uncertainty

[34] As a preliminary, the two rainfall-runoff models were run for all 913 gaged catchments. Here we split the available record period (1995–2005) into two periods, the first year being used each time for model warm-up (i.e., the models were actually calibrated and assessed over the 1996–2000 and 2001–2005 periods). Models were successively calibrated on the first period and tested in simulation mode on the second, and then calibrated on the second and tested on the first. Model performance could therefore be assessed each time and the parameter values on the two periods could be compared.

[35] Table 2 gives the models' median performance in calibration and verification for the two subperiods. In calibration mode, the medians of the Cr_Q criteria were approximately 0.82 for GR4J and 0.78 for TOPMO. In verification mode, the medians of the Cr_Q criteria were 0.78 for GR4J and 0.72 for TOPMO. This indicates that the performance of the two models decreases slightly when moving from calibration to verification. A discussion on the relative performances of the two models used is outside the scope of the paper. However, interestingly, the

Table 4. Correlation Between Model Parameters (Calibration Over the 1995–2005 Period)

Model		Parameters					
		X1	X2	X3	X4	X5	X6
TOPMO	X1	1.00					
	X2	0.09	1.00				
	X3	0.43	0.17	1.00			
	X4	−0.05	0.15	−0.12	1.00		
	X5	−0.11	−0.10	0.54	0.02	1.00	
	X6	0.22	−0.05	0.35	−0.07	0.22	1.00
GR4J	X1	1.00					
	X2	−0.20	1.00				
	X3	0.21	−0.39	1.00			
	X4	−0.29	0.02	−0.14	1.00		

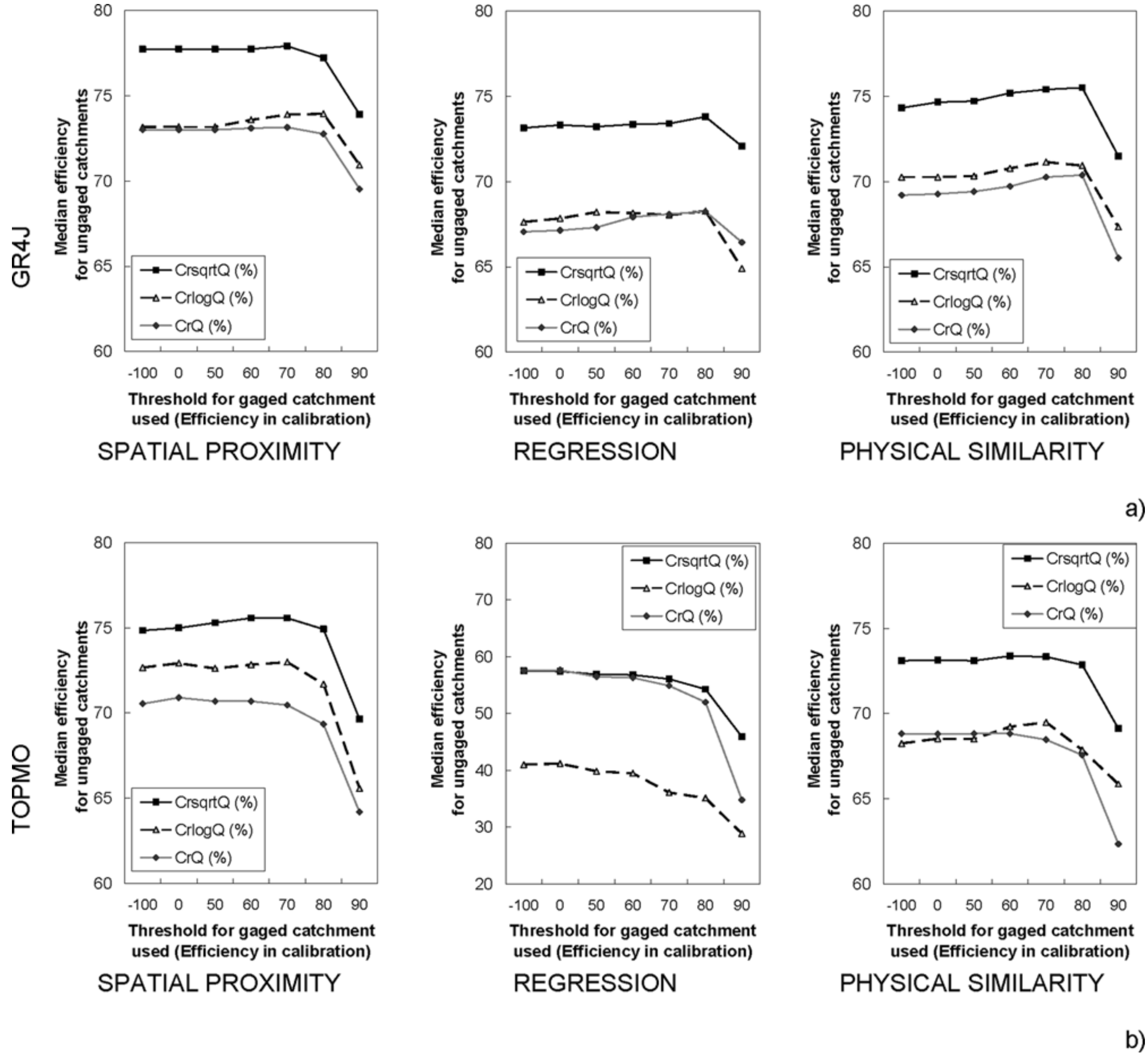


Figure 4. Impact of a poorly modeled gaged catchment on the efficiency of the three regionalization schemes tested for (a) GR4J and (b) TOPMO.

4-free-parameter GR4J model slightly outperforms the 6-free-parameter TOPMO model. These results are consistent with previous results [see, e.g., *Jakeman and Hornberger*, 1993; *Perrin et al.*, 2001] suggesting that a relatively low number of free parameters is sufficient to achieve reasonable streamflow simulations in many catchments. Figure 3 shows the spatial variation of the GR4J model efficiency in calibration. In the western catchments, model performance was significantly better than in the other parts of the country. Conversely, southern catchments are generally difficult to model since intense and spatially variable rainfall events make the streamflows vary strongly in time and amplitude. Note that, although not shown, similar spatial variations in calibration efficiency were obtained for TOPMO.

[36] We judged the stability of the model parameters by comparing their values obtained on the 1996–2000 and the 2001–2005 subperiods. Table 3 shows the correlation coefficient (R) of model parameters calibrated on the two independent subperiods. For GR4J, there was little scattering in the R values of the parameters, with the correlation coefficient up to 0.86, showing an acceptable degree of robustness. For TOPMO, the correlation coefficients exhibited a wider scattering, with values ranging from 0.29 to 0.73.

[37] We also analyzed the interdependence of model parameters to investigate possible overparameterization of the two rainfall-runoff models. Table 4 shows the matrices of correlation coefficients of calibrated model parameters over the catchment set. For GR4J, the interdependence of the calibrated parameters is weak, which probably stems from the parsimony of the model. For TOPMO, the interdependence of some parameters was more pronounced, particularly between X3 (the topographic index

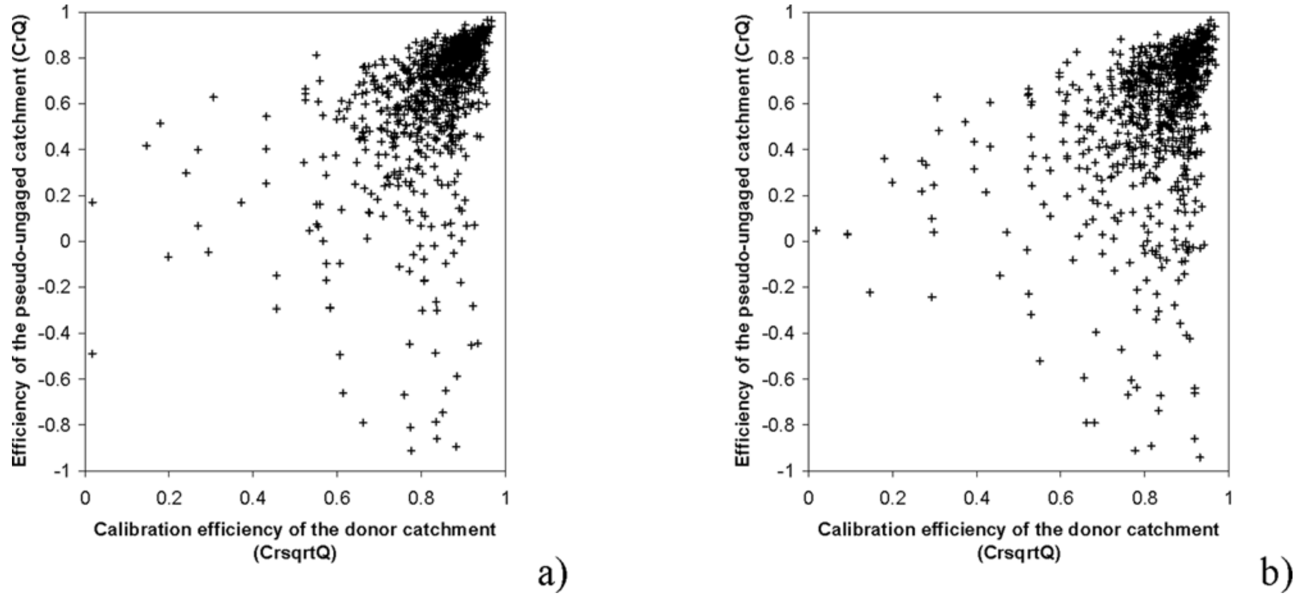


Figure 5. Relationship between the calibration efficiency of the model for donor gaged catchments and the efficiency of the model on pseudo-ungaged catchment. Case of GR4J regionalized following (a) the spatial proximity approach and (b) the physical similarity approach.

distribution parameter) and the others, but the coefficients of correlation were reasonably below 0.60. In the following sections, we will investigate the extent to which a reduction in the number of calibrated parameters tends to reduce the efficiency of certain regionalization schemes.

4.2. Optimal Settings for Each Regionalization Approach

[38] As seen in section 3, there are several options to be chosen within each regionalization approach. In this section, we investigate the impact of the chosen options on the efficiency of each regionalization scheme. Note that the optimal settings obtained here are applicable to our catchment set. Extrapolation to other catchment sets in other countries may not be warranted since these optimal settings depend on the choice of physical descriptors used for defining the physical similarity and the regression-based approaches, as well as on the density of the donor (gaged) basin network. When defining the optimal settings, several questions arise:

4.2.1. Should We Keep Poorly Modeled Catchments for Regionalization Purposes?

[39] The answer to this question of catchment outliers is not straightforward. On the one hand, we could suspect poorly modeled catchments of yielding highly uncertain model parameter values and on the other hand, poorly

Table 5. Impact of the Characteristics Used for the Physical Similarity Approach on the Efficiency of the Regionalized Model[a]

Number of Descriptors	Catchment Descriptors	GR4J (4 Donor Catchments) Median CrQ	GR4J (4 Donor Catchments) 25th–75th range percentile	TOPMO (7 Donor Catchments) Median CrQ	TOPMO (7 Donor Catchments) 25th–7th range percentile
6	*Ac, Zc, Sc, dc, fc, Em/Rm*	0.71	0.29	0.69	0.33
5	*Ac, Sc, dc, fc, Em/Rm*	0.71	0.31	0.69	0.35
5	*Zc, Sc, dc, fc, Em/Rm*	0.71	0.30	0.69	0.35
5	*Ac, Zc, dc, fc, Em/Rm*	0.71	0.31	0.69	0.33
4	*Sc, dc, fc, Em/Rm*	0.71	0.30	0.69	0.38
4	*Ac, Sc, fc, Em/Rm*	0.71	0.32	0.69	0.36
4	*Zc, dc, fc, Em/Rm*	0.71	0.29	0.68	0.35
3	*Sc, fc, Em/Rm*	0.71	0.30	0.69	0.38
3	*Zc, Sc, Em/Rm*	0.71	0.30	0.69	0.36
3	*Zc, fc, Em/Rm*	0.70	0.31	0.69	0.36
2	*Sc, Em/Rm*	0.69	0.30	0.67	0.38
2	*dc, Em/Rm*	0.69	0.29	0.67	0.38
2	*Zc, Em/Rm*	0.69	0.32	0.67	0.38
1	*Em/Rm*	0.66	0.35	0.66	0.39
1	*Zc*	0.66	0.35	0.65	0.37
1	*dc*	0.65	0.35	0.65	0.35
0	random sampling of gaged catchments	0.59	0.24	0.60	0.31

[a]Median CrQ and range of the 25th and 75th percentiles. A_c: catchment area, S_c: mean slope, Z_c: median altitude, d_c: river network density, f_c: fraction of forest cover, E_m/R_m: aridity index.

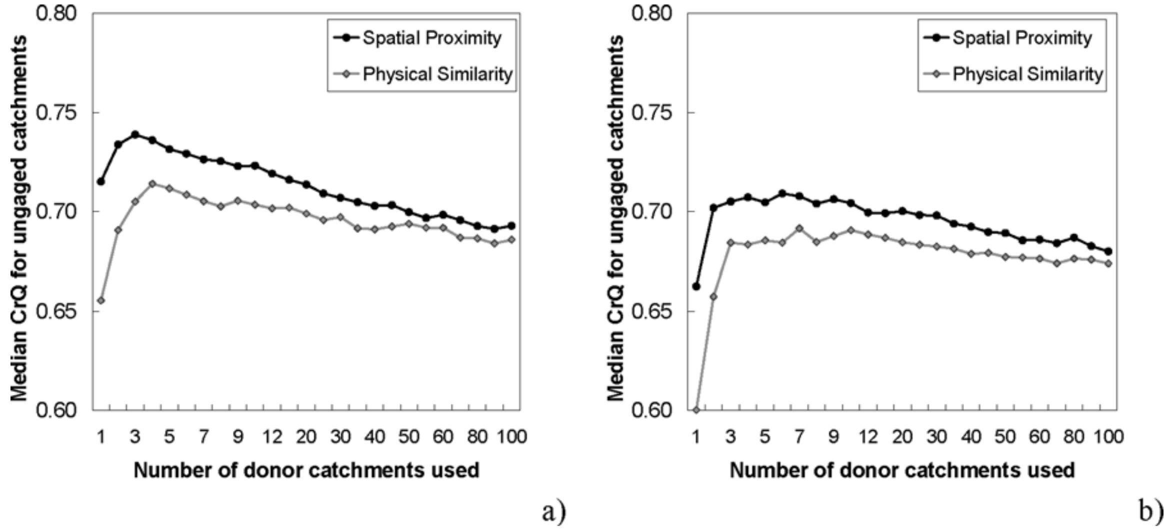


Figure 6. Impact on model efficiency of the number of gaged basins used for ungaged basin prediction. Spatial proximity and the similarity regionalization schemes for (a) GR4J and (b) TOPMO.

modeled catchments may introduce a diversity that could prove beneficial for modeling ungaged catchments. To determine whether poorly modeled catchments should be kept in our comparative study, we simply used a threshold on model efficiency (in calibration mode) under which a donor (gaged) catchment was not used to predict streamflows on ungaged sites. However, note that all catchments, whether poorly or well modeled, were considered in turn to be ungaged. Results for the three regionalization schemes are plotted in Figure 4. It can be seen that an intermediate value for the threshold (approximately 0.70) on the criterion Cr_{sqrtQ} in calibration yields the best regionalization results. A threshold lower than 0.70 does not greatly affect the performance of the model for ungaged catchments, whereas imposing a threshold of 0.90 for Cr_{sqrtQ} is detrimental for the regionalization

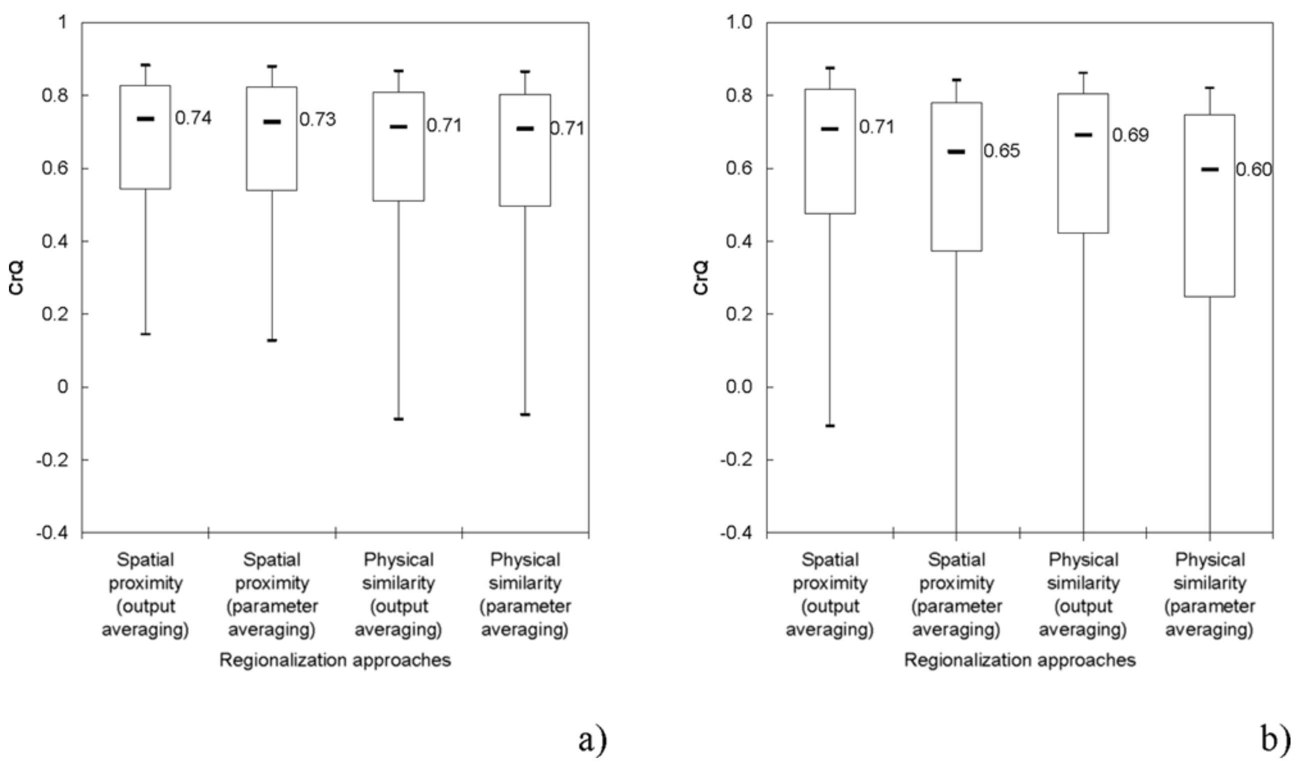


Figure 7. Comparison of the model output averaging option and the model parameter values averaging options for the spatial proximity and the physical similarity approaches. (a) GR4J and (b) TOPMO. The boxes are delimited by the 25th and 75th percentiles, the median is marked with a thick line and the whiskers are delimited by the 10th and 90th percentiles.

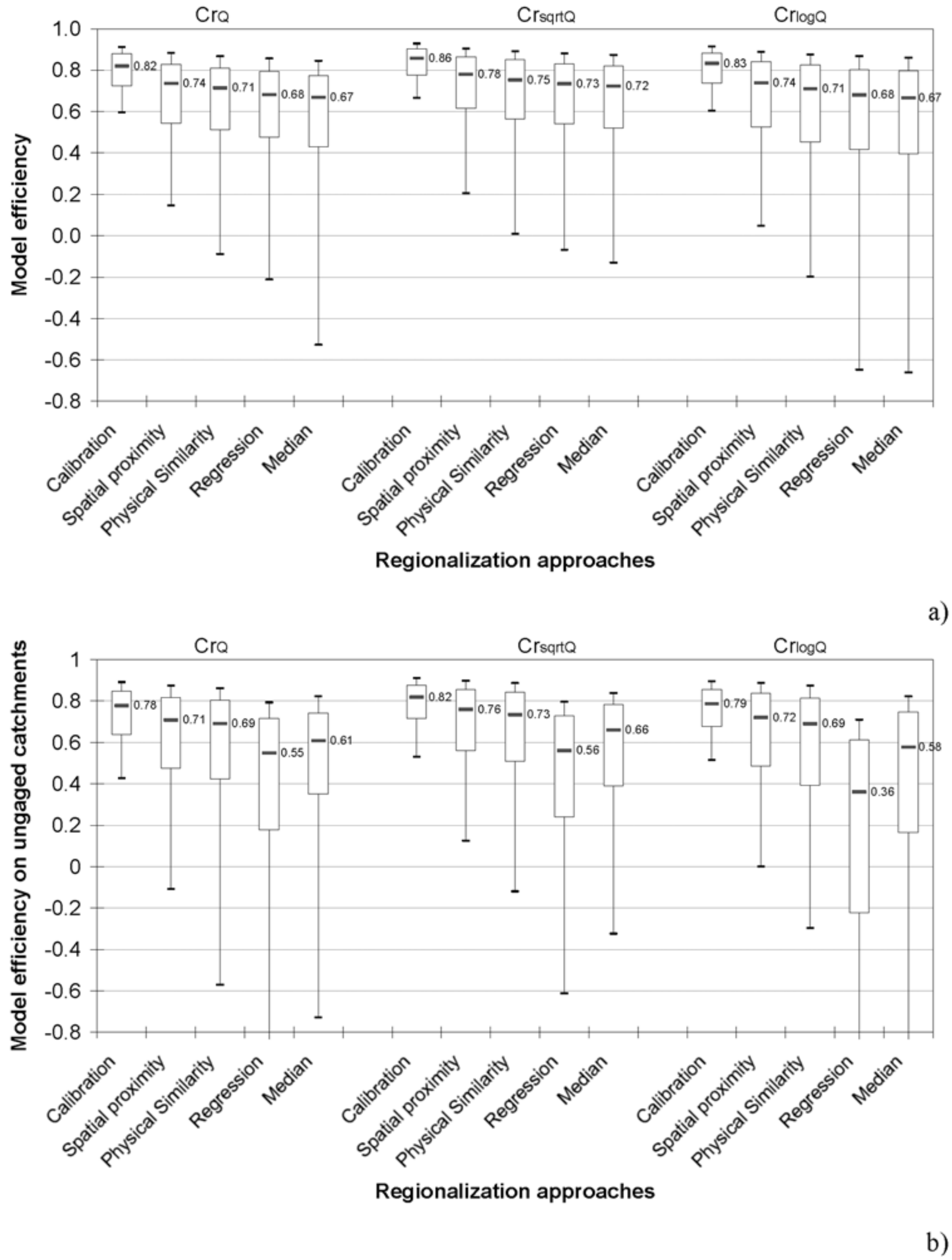


Figure 8. Comparison of GR4J (a) and TOPMO (b) efficiencies on ungaged catchments using several regionalization schemes. The boxes are delimited by the 25th and 75th percentiles, the median is marked with a thick line and the whiskers are delimited by the 10th and 90th percentiles.

studies. This is probably attributable to the number and the particular location of the gaged catchments considered: for a threshold of about 0.90, only 253 (using the GR4J model) and 150 (using the TOPMO model) gaged catchments out of the 912-catchment set were available for use; these catchments being mainly located in Brittany (western France; see Figure 3). To shed more light on this issue, Figure 5 shows the relationship between the calibration efficiency of the model on the donor gaged catchment and the efficiency of the model on the pseudo-ungaged catchment, for the particular case of one single donor. Results suggest that using a well-modeled catchment as donor does not warrant a good level of efficiency of the model for the pseudo-ungaged catchment. However conversely, if only the worst modeled catchments are used as donors, the performances of the regionalization approaches are clearly affected, due probably to particular values of model parameters when calibrated over those catchments.

[40] Hereafter we will use a value of 0.70 for the threshold on the criterion obtained in calibration for the gaged catchments using both the GR4J model and the TOPMO model. If a gaged catchment obtains an efficiency rating lower than this value in calibration, it will not be considered as a potential donor for hydrological information transfer, irrespective of which regionalization approach is chosen.

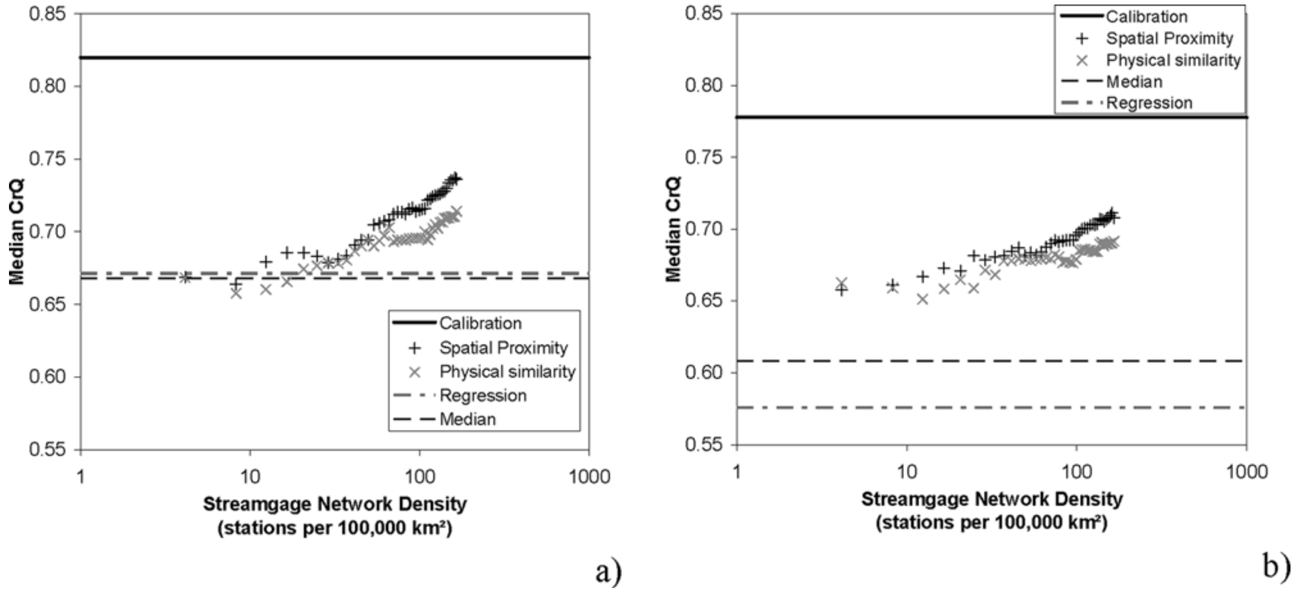


Figure 9. Impact of a reduction of the streamgage network density of possible donor catchments on the performance of the spatial proximity approach and the physical similarity approach. Case of GR4J (a) and TOPMO (b).

4.2.2. For the Physical Similarity Approach, Which Catchment Descriptors are the Most Informative?

[41] Table 5 shows the range of model efficiencies obtained with the similarity approach when choosing different combinations of catchment characteristics to assess the similarity between two catchments. As a reference, we also present the efficiency of a blind combination, where the donor catchments were selected randomly. We present only the results for the Cr_Q criterion since the other criteria behave similarly. Table 5 shows that three out of the six catchment descriptors used are sufficient to obtain optimal efficiencies. Increasing the number of descriptors does not increase model performance, which is probably due to the redundant information among the descriptors used in this study. This interdependence of some descriptors also yields relatively homogeneous results whatever the combination used, even though the aridity index (E_m/R_m) would appear to be the most relevant variable. Hereafter, we will present the results obtained with a combination of only three catchment descriptors, namely the mean slope (S_c), the fraction of forest cover (f_c) and the aridity index (E_m/R_m). This combination offers the best solution for the two selected rainfall-runoff models, with the median Cr_Q value equal to 0.71 for GR4J and 0.69 for TOPMO. This level of performance is significantly greater than that of the blind combination, which yields median Cr_Q value around 0.60 for both models.

4.2.3. For the Physical Similarity Approach and the Spatial Proximity Approach, How Many Catchments are Needed as Donors?

[42] In order to investigate this issue, we tested an increasing number of donor catchments from one to 100. Figure 6 shows that for regionalization based on spatial proximity and physical similarity, only a few donor catchments (approximately five) were required. Choosing only one donor catchment was detrimental for regionalization purposes: a larger number of donor catchments allows avoiding strong errors in streamflow simulation by smoothing the response with other sources. However, the efficiency of the regionalized model decreased almost linearly when increasing the number of donor catchments (above five) that are further and further from the ungaged catchment. Last, note that the two models behaved similarly in this regard but the optimal number of donor catchments was slightly different for the two models: GR4J required four donor catchments whereas TOPMO required seven donor catchments to reach its optimal efficiency. This difference is not significant enough to draw any conclusion but these optimal settings will be kept hereafter.

4.2.4. For the Similarity Approach and the Spatial Proximity Approach, is it Better to Average Model Parameter Values (Equation 1) or Model Outputs (Equation 2)?

[43] For both models, Figure 7 shows plots of the distribution of the Cr_Q values obtained when averaging model outputs against the output obtained when averaging model parameters for the spatial proximity regionalization scheme and the similarity regionalization scheme. For GR4J, the efficiencies of the two options are similar, even if the model output averaging option is slightly more efficient, with an increment around 0.01 for the median of the Cr_Q. For the less efficient TOPMO, the benefit of averaging model outputs rather than model parameters is more pronounced since it yields an increment above 0.06 of the median Cr_Q. This could be caused by the greater interdependence of the TOPMO parameters, which is more or less damped out when individually averaging model parameters. Hereafter, we will only consider averaging on model output since it is more efficient for the TOPMO model and intuitively more satisfying since it uses the hydrological behavior of a donor catchment as characterized by its entire parameter set, while the parameter averaging method takes each parameter at its face value, thus neglecting the unavoidable interactions between parameters.

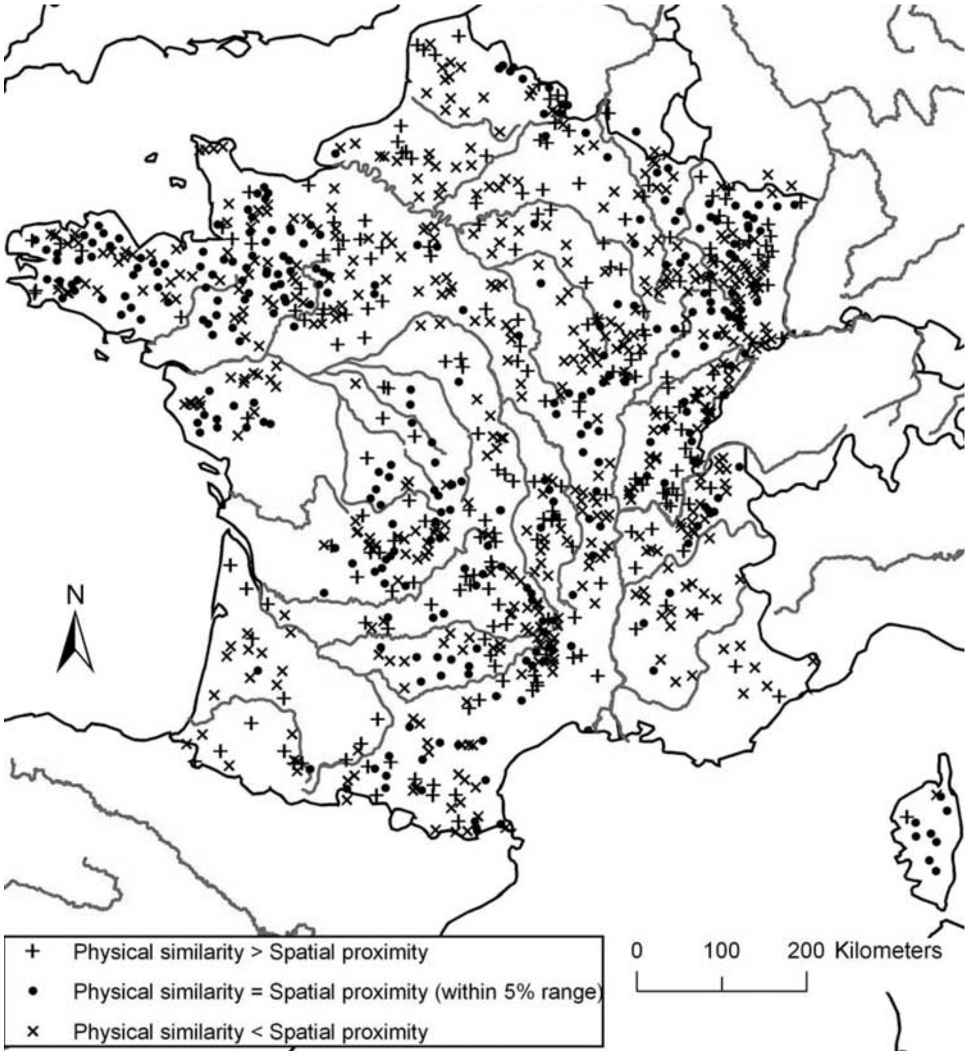


Figure 10. Comparison of the performance of two regionalization approaches: physical similarity and spatial proximity. GR4J.

[44] Having considered the above four questions, the optimal settings retained for each regionalization scheme are summarized below:

[45] • For the regression-based approach, we keep only the catchments with calibration efficiency (Cr_{sqrtQ}) above 0.70 to derive the regression equations. All catchment descriptors are used in these equations, even if they are not all warranted by significant Student t ratio.

[46] • For the spatial proximity approach, a selection of four (for GR4J) or seven (for TOPMO) donor gaged catchments based on the distances between catchment centroids and the model output averaging option are preferred. Only catchments with calibration efficiency (Cr_{sqrtQ}) above 0.70 are considered as possible donors.

[47] • For the physical similarity approach, the model output averaging option is also preferred. We consider three catchment descriptors, namely the mean slope (S_c), the fraction of forest cover (f_c) and the aridity index (E_m/R_m). For each catchment descriptor, a ranking of the possible donor catchments is performed. Then, the mean of the ranks is used to select four (for GR4J) or seven (for TOPMO) gaged catchments that are the most similar to the ungaged one. Only catchments with calibration efficiency (Cr_{sqrtQ}) above 0.70 are considered as possible donors.

4.3. Comparison of the Different Regionalization Schemes in Terms of Model Efficiency

4.3.1. Comparison Over the 913-Catchment Set

[48] Using the optimal settings selected earlier, we compared the performance of the three different approaches for regionalization. Two reference efficiencies were also considered: the model performance obtained in calibration and that obtained using the same median parameter set for the 913 catchments (as the simplest possible form of regionalization). The calibration results represent the upper (practically unreachable) limit of the regionalization method whereas the results obtained with the median set of parameters were expected to be the worst regionalization result. Thus the results obtained with the three regionalization schemes tested were expected to lie between the two extremes.

[49] Figure 8 summarizes the results by showing the distributions of the model efficiencies given by the three assessment criteria. The differences found between the three regionalization methods were clear but not very large. For

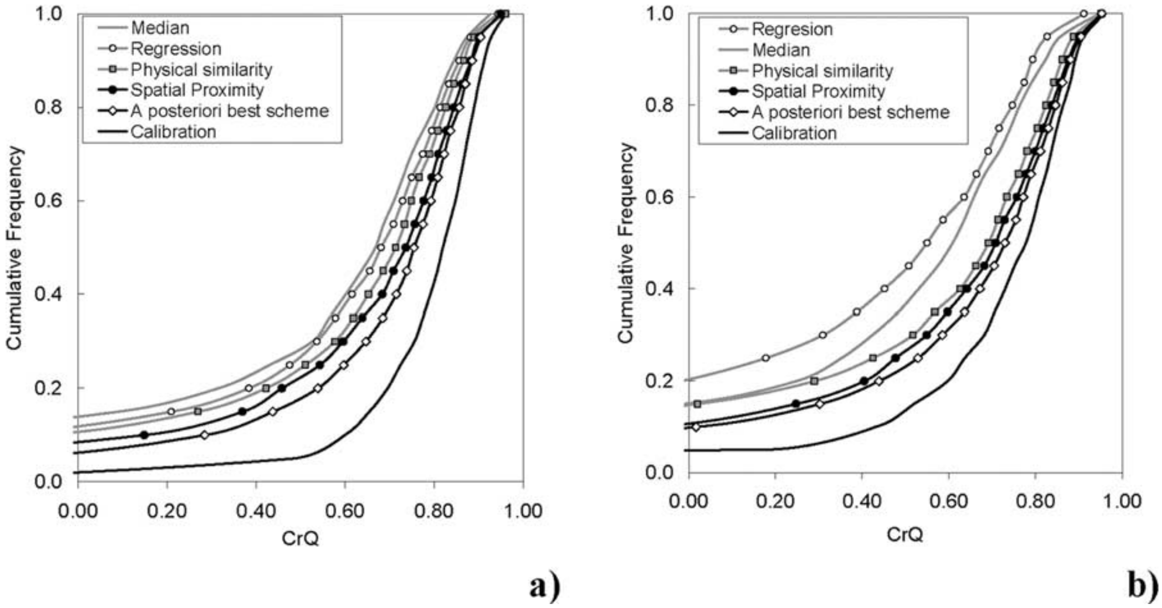


Figure 11. Distribution of CrQ values obtained by different regionalization schemes, including the best scheme a posteriori. GR4J (a) and TOPMO (b).

all three efficiency criteria, the ranking of the three approaches was the same on our selected catchment set. Spatial proximity offered the best regionalization solution (with values of the median Cr_Q equal to 0.73 and 0.71 for GR4J and TOPMO, respectively). Regression was the worst (the values of the median Cr_Q equal to 0.68 and 0.55, respectively), with results close to the primitive median parameter solution (having values of the median Cr_Q equal to 0.67 and 0.61, respectively). The physical similarity approach was inbetween the two (with values of the median Cr_Q equal to 0.71 and 0.69, respectively). Note also that the superiority of the spatial proximity approach over the other regionalization methods is even more pronounced when looking at the lower percentiles, indicating that this approach is much more robust than the others.

[50] Comparing the performance of the two models for each regionalization scheme also provides interesting insights. The two models were quite close in efficiency for the physical similarity approach and also for the spatial proximity approach. Conversely, the model performances differed greatly when using the regression approach, perhaps because the physical similarity and spatial proximity approaches used model output averaging and thereby facilitated compensation for possible overparameterization of TOPMO compared to GR4J.

[51] For the spatial proximity and the physical similarity approaches, nested catchments are likely to play an important role, as demonstrated by *Merz and Blöschl* [2004]. Over the 913 catchments used in this study, there is a relatively high degree of nesting, with more than 500 catchments that have upstream and/or downstream neighbors. To test the impact of these nested catchments on the performance of the regionalization schemes, we excluded the nested catchments as possible donors. Results showed that the performances are only slightly affected, since the loss in median Cr_Q values is 0.01 for the two models and for the two regionalization approaches.

4.3.2. Impact of the Streamgage Network Density on the Relative Efficiencies of the Regionalization Methods

[52] The relative levels of efficiency reached for the spatial proximity and the physical similarity approaches are likely to depend on the density of gaging stations, and the leading position of the spatial proximity may fade away when this density decreases. To assess the impact of network density, we progressively decreased the density of possible donor catchments being used for each ungaged catchment, the overall assessment of the regionalization schemes still being made on the 913 catchments. Figure 9 presents the evolution of the median of models' efficiencies when decreasing network density. Interestingly, the two approaches become similar in terms of efficiency when the streamgage network is around 60 stations per 100,000 km^2, i.e., when reducing by one third the actual density of catchments used. Below this density, results for the two approaches become virtually the same and it was not possible to decide which approach is the most appropriate.

4.3.3. Possible Complementarity of the Regionalization Approaches

[53] To examine whether there are spatial patterns in the performance of the regionalization methods, Figure 10 shows, for each catchment, the best regionalization solution (as between physical similarity and spatial proximity) for the GR4J model. Interestingly, the spatial proximity approach did not systematically outperform the physical similarity approach, suggesting that these two approaches could be used in a complementary way.

[54] To investigate the potential of combining those two regionalization approaches, for each pseudo-ungauged catchment we retained the best performance in simulation mode provided by the physical similarity and the spatial proximity approaches (ignoring the regression-based approach as that did not improve the overall performance). Figure 11 shows the cumulative frequency distributions of Cr_Q values obtained by this best a posteriori multiregional-

ization choice together with those of the other approaches. There is a significant gap between the best multiregionalization scheme and the others, meaning that one approach contains specific information that can be satisfactorily complemented by the other. However, further research is needed to determine which rules would be needed to select a priori the multiregionalization approach that would yield the best model performance. From Figure 10, it can only be deduced that these rules are not straightforward. However, an extension of the model output averaging option could be to combine flow simulations obtained by the spatial proximity and the physical similarity approaches as a new variant of these regionalization schemes.

5. Summary and Conclusion

[55] The aim of this study was to compare three classical options to regionalizing rainfall-runoff models for use on ungaged basins: the regression-based approach, an approach based on physical similarity and the spatial proximity approach. Our results show that spatial proximity offers the best regionalization solution, regression providing the worst of the three, with results not much better than those obtained with the median parameter solution, the physical similarity approach being intermediary. Thus these results are consistent with recent Austrian case studies [*Merz and Blöschl*, 2004; *Parajka et al.*, 2005] in which the spatial proximity approach provided the most valuable option. However, when comparing our results with calibration efficiencies, it is obvious that there is still considerable room for progress on the regionalization issue, comparable to that reported in other studies on large sets of catchments [*Merz and Blöschl*, 2004; *McIntyre et al.*, 2005; *Young*, 2006].

[56] Compared to the studies by *McIntyre et al.* [2005] and *Young* [2006], the failure of the regionalization approaches based on catchment descriptors (regression and physical similarity) may be attributable to the lack of a key physical descriptor, namely, soil type, which is not available at the national level in France and therefore not considered here. However, it may also be that on our large and much more diverse data set, physical descriptors are not relevant enough to explain catchment behavior. Note also that the dense network of basins used in this study may have favored the spatial proximity approach. Therefore even if we now strongly recommend the spatial proximity approach over the physical similarity one, transposing this approach to regions where gaged basins are sparse might well lead to less favorable results than those reported here, but this should be investigated further.

[57] We consider that given the size of our data set and the fact that we used two different rainfall-runoff models, our results provide general insights on the regionalization issue. As far as perspectives for improvement are concerned, it seems that substantial progress could be made if we knew how to use the different regionalization approaches in a complementary fashion, i.e., in a multiregionalization approach.

[58] **Acknowledgments.** The authors thank Murugesu Sivapalan, Ralf Merz, Neil R. Viney and the two anonymous reviewers for their constructive suggestions, which helped to improve the text. This research is based on a very large hydrometeorological data set, and we would like to acknowledge: The contribution of Météo France in providing the SAFRAN pluviographic archive (special thanks go to Joël Noilhan and Laurent Franchisteguy); The contribution of the SCHAPI in providing access to the HYDRO streamflow archive (special thanks go to Laurent Cobos); The work of Jean-louis Rosique who digitized most of the 913 catchment boundaries.

References

Acreman, M. C., and C. D. Sinclair (1986), Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland, *J. Hydrol.*, *84*, 365–380.

Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*, 703–710.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29.

Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*, 43–69.

Burn, D. H., and D. B. Boorman (1993), Estimation of hydrological parameters at ungauged catchments, *J. Hydrol.*, *143*, 429–454.

Edijatno, N. Nascimento, X. Yang, Z. Makhlouf, and C. Michel (1999), GR3J: A daily watershed model with three free parameters, *Hydrol. Sci. J.*, *44*, 263–278.

Egbuniwe, N., and D. K. Todd (1976), Application of the stanford watershed model to Nigerian watersheds, *Water Resour. Bull.*, *12*, 449–460.

Franchini, M., J. Wendling, C. Obled, and E. Todini (1996), Physical interpretation and sensitivity analysis of the TOPMODEL, *J. Hydrol.*, *175*, 293–338.

Goswami, M., and K. M. O'Connor (2006), Flow simulation in an ungauged basin: an alternative approach to parameterization of a conceptual model using regional data, in *Large sample basin experiments for hydrological model parameterisation*, edited by V. Andréassian et al., pp. 149–158, IAHS publication n°307, Wallingford.

Goswami, M., K. M. O'Connor, and K. P. Bhattarai (2007), Development of regionalization procedures using a multi-model approach for flow simulation on an ungauged catchment, *J. Hydrol.*, *333*, 517–531.

Gottschalk, L., J. L. Jensen, D. Lundquist, R. Solantie, and A. Tollan (1979), Hydrologic regions in the Nordic countries, *Nord. Hydrol.*, *10*, 273–286.

Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, *292*, 281–295.

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

James, L. D. (1972), Hydrologic modeling, parameter estimation, and watershed characteristics, *J. Hydrol.*, *17*, 283–307.

Kay, A. L., D. A. Jones, S. M. Crooks, A. Calver, and N. S. Reynard (2006), A comparison of three approaches to spatial generalization of rainfall-runoff models, *Hydrol. Processes*, *20*, 3953–3973.

Magette, W. L., V. O. Shanholtz, and J. C. Carr (1976), Estimating selected parameters for the Kentucky watershed model from watershed characteristics, *Water Resour. Res.*, *12*, 472–476.

Mathevet, T. (2005), Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et intercomparaison de modèles sur un large échantillon de bassins versants, PhD thesis, Cemagref, Antony.

McIntyre, N., H. Lee, H. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2005WR004289.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*, 95–123.

Merz, R., G. Blöschl, and J. Parajka (2006), Regionalisation methods in rainfall-runoff modelling using large samples, in *Large sample basin experiments for hydrological model parameterisation*, edited by V. Andréassian et al., IAHS publication n°307, Wallingford.

Michel, C., C. Perrin, and V. Andréassian (2003), The exponential store: A correct formulation for rainfall–runoff modelling, *Hydrol. Sci. J.*, *48*, 109–124.

Nascimento, N. O. (1995), Appréciation à l'aide d'un modèle empirique des effets d'action anthropiques sur la relation pluie-débit à l'échelle du bassin versant, PhD thesis, 550 pp, CERGRENE/ENPC, Paris.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models Part one: A discussion of principles., *J. Hydrol.*, *27*, 282–290.

Nathan, R. J., and T. A. McMahon (1990), Identification of homogeneous regions for the purposes of regionalisation, *J. Hydrol.*, *121*, 217–238.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne (2005), Which potential evapotranspiration input for a lumped rainfall-runoff model? - Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, *J. Hydrol.*, *303*, 290–306.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006a), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *J. Hydrol.*, *320*, 62–83.

Oudin, L., V. Andréassian, T. Mathevet, C. Perrin, and C. Michel (2006b), Dynamic averaging of rainfall-runoff model simulations from complementary model parameterization, *Water Resour. Res.*, *42*(7), W07410, doi:10.1029/2005WR004636.

Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, *9*, 157–171.

Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, *43*, W06406, doi:10.1029/2006WR005271.

Pardé, M. (1933), Fleuves et Rivières, 224 pp., Paris.

Perrin, C. (2000), Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative, PhD thesis, 530 pp, INP/Cemagref Antony, Grenoble, France.

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

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*, 275–289.

Perrin, C., L. Oudin, V. Andréassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*, 131–151.

Rojas-Serna, C. (2004), Quelle connaissance hydrométrique minimale pour définir les paramètres d'un modèle pluie-débit? PhD thesis, 319 pp, ENGREF/Cemagref, Paris.

Vandewiele, G. L., C. Y. Xu, and W. Huybrecht (1991), Regionalisation of physically-based water balance models in Belgium: Application to ungauged catchments, *Water Resour. Manage.*, *5*, 199–208.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, *181*, 23–48.

Young, A. R. (2006), Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, *320*, 155–172.

V. Andréassian, N. Le Moine, C. Michel, and C. Perrin, Cemagref, Hydrosystems and Bioprocesses Research Unit, Parc de Tourvoie, BP 44, 92163 Antony Cedex, France.

L. Oudin, UPMC Univ Paris 06, UMR 7619, Sisyphe, F-45005, Paris, France. (ludovic.oudin@ccr.jussieu.fr)