How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments.

Claudia Rojas-Serna, Laure Lebecherel, Charles Perrin, Vazken Andréassian, Ludovic Oudin

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

How should a rainfall-runoff model be parameterized in an almost ungaged catchment? A methodology tested on 609 catchments

Claudia Rojas-Serna[1], Laure Lebecherel[2], Charles Perrin[2]*, Vazken Andréassian[2], Ludovic Oudin[3]

[1] Universidad Autónoma Metropolitana, División de Ciencias Básicas e Ingeniería, México, Mexico

[2] Irstea, Hydrosystems and Bioprocesses Research Unit (HBAN), Antony, France

[3] UMR Metis, Université Pierre et Marie Curie, Paris, France

* Corresponding author, e-mail: charles.perrin@irstea.fr; Fax: +33 1 40 96 61 99

Key points

- A methodology to use point flow measurements for parameter estimation is presented
- Tests were made on a set of catchments with various flow conditions
- Better efficiency than classical regionalization approaches with only a few flow measurements

Abstract

This paper examines catchments that are *almost* ungaged, i.e. catchments for which only a small number of point flow measurements are available. In these catchments, hydrologists may still need to simulate continuous streamflow time series using a rainfall-runoff model, and the methodology presented here allows using few point measurements for model parameterization. The method combines regional information (parameter sets of neighboring gaged stations) and local information (contributed by the point measurements) within a framework where the relative weight of each source of information is made dependent on the number of point measurements available. This approach is tested with two different hydrological models on a set of 609 catchments in France. The results show that on average a few flow measurements can significantly improve the simulation efficiency, and that ten measurements can reduce the performance gap between the gaged and ungaged situations by more than 50%. Model parameters tend to come closer to the values obtained by calibration in fully gaged conditions as the number of point flow measurements increases.

Keywords: Rainfall-runoff modeling; Regionalization; Point flow measurements; Ungaged catchment; Neighborhood; Parameter estimation

1 Introduction

1.1 Parameter estimation on entirely ungaged catchments

Prediction in ungaged basins (PUB) has been one major focus of the hydrological community in the past decade [*Sivapalan*, 2003] but still remains a great challenge. A full review of ungaged basin research is not within the scope of this article and readers are referred to the recent reviews by *Blöschl et al.* [2013] and *Hrachowitz et al.* [2013].

Here, we only mention that the most common parameter estimation techniques to transfer information from gaged (donor) to ungaged (target) catchments are based on: (1) regression equations linking parameters to physical and climate catchment characteristics (regressions to be calibrated on gaged catchments), (2) transfer of parameter sets obtained in gaged catchments that are similar in terms of physical and climate characteristics to the target catchment, (3) transfer of parameter sets obtained in geographically close catchments. Many variants exist, including procedures of regional model calibration [*Castiglioni et al.*, 2010; *Fernandez et al.*, 2000; *Hundecha et al.*, 2008; *Lombardi et al.*, 2012; *Parajka et al.*, 2007].

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

These different approaches have been compared in various contexts, sometimes producing contradictory results. *Parajka et al.* [2013] made a cross-analysis of 34 past regionalization studies, showing that climate conditions or network density can explain differences between results. In dense network conditions, spatial proximity approaches are often those that perform the best.

1.2 Point streamflow measurements are ubiquitous

To cope with the difficulties of estimating parameters in ungaged catchments, using complementary or soft data, i.e. additional measurements or information on the catchment, was advocated by some authors [*Fenicia et al.*, 2008; *Seibert and McDonnell*, 2002; *Winsemius et al.*, 2009]. Among these complementary data, short time series or point streamflow measurements are increasingly recognized as a valuable source of information for model parameterization [see e.g. *Tada and Beven*, 2012]. There are indeed many locations in the world where it is difficult to maintain long-term flow gauging stations and where only short series are available [see e.g. *Bhatt and Tiwari*, 2008; *Kim and Kaluarachchi*, 2009; *Konz et al.*, 2007]. Also, when a hydrological question is raised for an ungaged river, practicing hydrologists may not have the time and resources to install a perennial gauging station, but they will generally have the opportunity to take a small number of flow measurements.

One can mention a few examples of such situations reported in the literature. *Hughes et al.* [2014] made weekly flow measurements on a small stream in South Africa over an 18-month period and investigated how these data can help better constrain the parameters of the Pitman hydrological model. *Temnerud et al.* [2007] carried out point flow measurements at 66 sites within a 78-km^2 catchment in northern Sweden during a low-flow period to investigate the role of spatial patterns in water quality issues. In France, systematic point flow measurement campaigns have been coordinated by the Rhine-Meuse Water Agency over the last two decades to improve the knowledge of low flows [*Corbonnois et al.*, 1999; *Decloux and Sary*, 1991; *Drogue and Plasse*, 2014; *François and Sary*, 1990; 1994; *Plasse et al.*, 2014]. These point flow measurements may be useful for a number of objectives, such as low-flow estimation [*Catalogne et al.*, 2014; *Chopart and Sauquet*, 2008; *Eng and Milly*, 2007; *Goodwin and Young*, 2007; *Laaha and Blöschl*, 2005; *Oberlin et al.*, 1973].

Although the hydrologist needing to calibrate a model will not consider these catchments with short times series or point flow measurements as *properly* gaged, they are not strictly speaking ungaged. In this paper, we will refer to them as *almost ungaged*.

1.3 Sensitivity of model performance to flow data availability

For gaged or almost ungaged catchments, the robustness of the parameter sets identified clearly depends on the information content of flows available for calibration [*Wagener et al.*, 2003]. When information is lacking, the *mathematical optimum* identified during calibration will be potentially different from the *hydrological optimum* (i.e. the parameter set that can reproduce catchment behavior over the long term), with possible problems of over-calibration on the data at hand [see *Andréassian et al.*, 2012, for a more complete discussion on this issue]. Hence, increasing the length of the flow series available for calibration generally tends to produce more robust parameter sets, and several authors advised using series covering 5–8 years to calibrate models [see e.g. *Anctil et al.*, 2004; *Yapo et al.*, 1996]. Similarly, Gill et al. [2007] showed that increasing the percentage of missing data in a calibration series tends to produce less robust models. Indeed, longer time series generally encompass a larger variety of hydrological conditions, which makes the series more informative for calibration.

However, a number of studies tend to indicate that shorter time series may also provide valuable information. *Brath et al.* [2004] reported tests using calibration periods ranging from 1.5 to 12 months. The best results in validation were obtained using the parameters calibrated in the 12-month period, but acceptable results could be obtained with 3 months of data. Similarly, *Melsen et al.* [2014] concluded that 5 months of data are sufficient for parameter optimization to obtain good results on the full observation period on a Swiss catchment when applying a two-parameter model. Using discontinuous series, *Kim and Kaluarachchi* [2009] tested the sensitivity of a water balance model to decreasing data availability. They showed that short data series could produce parameter sets which can be effectively transposed in time as long as some parts of the hydrograph (especially high and low flows) are included in the series.

The sensitivity of model performance to flow availability was also assessed by *Perrin et al.* [2007]. They used from 10 to 1,000 flow data randomly sampled out of long series to calibrate two hydrological models on 12 US catchments. They showed that the optimized parameter values became stable for the two models when 350 flow data were available for calibration (with a significant drop in performance when fewer data were used).

1.4 Making the most of limited flow data

The problem of parameter estimation in almost ungaged catchments has received increasing attention over the last few years. Several authors suggested making explicit use of point flow measurements. *Montanari and Toth* [2007] proposed a practical approach to calibrate rainfall-runoff models with sparse data, using the Whittle estimator as a likelihood function and calibrating the model in the spectral domain. *Seibert and Beven* [2009] used point flow measurements to constrain the choice of model parameters. They concluded that "surprisingly little runoff data was necessary to identify model parameterizations that provided good results for the 'ungaged' test periods. These results indicated that a few runoff measurements can contain much of the information content of continuous runoff time series." *Tada and Beven* [2012] tested various optimization options of the TOPMODEL parameters on three Japanese catchments using short continuous periods of 4, 8, …, 512 days selected within a 10-year period. They showed that considering an ensemble of acceptable parameter sets could improve the results obtained by classic optimization when calibration time series are shorter than 1 year. *Singh and Bardossy* [2012] also proposed an approach to identify robust parameter sets when only short calibration periods are available. Using critical events identified in a series with a depth function, they showed that calibration on events representing 6–7% of a 10-year time series provided similar results to calibration on the whole time series and better results than random selection of events.

1.5 Using regional information in parameter estimation on almost ungaged catchments

Although the previous approaches concentrate on extracting information from the available flow data only, the knowledge on parameter values gained from gaged catchments was recognized early as valuable information even in the context of calibration procedures. For example, Koren et al. [2003] found it beneficial to use the regionalization relationships established for ungaged catchments as initial estimates of parameter values in the calibration process. Kuzmin et al. [2008] also underlined the advantage of starting calibration with already informative initial values and then improving these a priori estimates during the calibration process. These authors mention that searching in the vicinity of the starting values using a local search algorithm yields better results than using a global search algorithm, especially where calibration data are lacking.

These results encouraged hydrologists to develop methods for exploiting regional information in the case of limited flow information, to better constrain parameter estimation.

One solution is to limit the exploration of the parameter space during calibration. *Perrin et al.* [2008] showed that the drop in performance caused by shortening calibration time series was effectively attenuated by searching for the optimal parameter set in pre-sampled parameter space, i.e. within a collection of parameter sets previously obtained in other gaged catchments. The authors showed that this approach outperforms the classic calibration approaches when only short-flow time series are available. The results of *Andréassian et al.* [2014] corroborate these findings: they found that choosing parameters within a short-list of 27 parameter sets was more robust than making full calibration of the GR4J model when less than 1 year of flow data is available.

From a different perspective, *Seibert and McDonnell* [2013] proposed an approach combining the use of point flow measurements and soft data (user-defined fuzzy rules of acceptance of groundwater contributions). They used data collected over a 3-month period and showed that a single event or ten observations during high flows provided the same information as the continuous 3 months. *Winsemius et al.* [2009] also proposed a framework to integrate hard and soft data to constrain the estimation of parameters. Their results confirm the potential of their method to be used for almost ungaged catchments.

Another option is to combine parameters estimated by regionalization and point flow measurement. Among the first attempts, *Rojas-Serna et al.* [2006] presented a method merging the use of regionalized parameter values and optimization to parameterize a model when only a few point measurements were available. More recently, *Viviroli and Seibert* [2015] proposed a framework to improve parameter constraints with point flow measurements. On a set of catchments in Switzerland, they showed that even a few flow measurements help constrain parameters and improve model performance over purely regionalized parameters.

The above studies suggest that regional sources of information can be useful for parameter estimation even in the classic case of model calibration at a long-term gauging station. In almost ungaged stations, on which this paper focuses, starting from regional hydrological knowledge appears promising.

1.6 Scope of the paper

The above review suggests that parameters can be estimated using a limited number of flow observations (not necessarily contiguous but possibly spread out in time), but that classic optimization algorithms are insufficient for model parameterization when only a few flow data are available for calibration. Exploiting the prior knowledge gained at a regional level (in physically or spatially neighboring gauging catchments) is a valid alternative for parameter identification.

This paper presents a new parameter estimation approach for almost ungaged catchments, which specifically combines regional information transferred from neighboring gaged catchments with local information contributed by a limited number of flow measurements. The framework we present in this paper intends to make use of these measurements for the identification of hydrological model parameters, thus making a bridge between the fully ungaged and gaged cases, as illustrated in Figure 1.

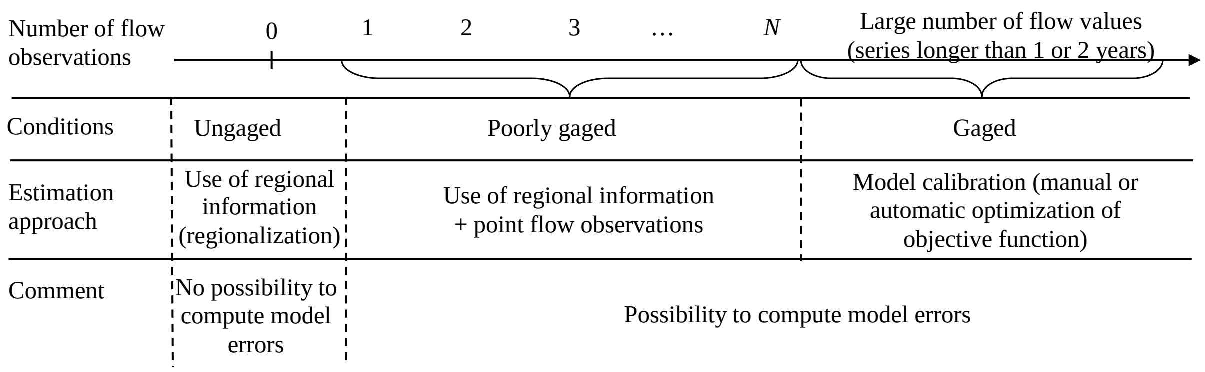

Figure 1. Comparison of the ungaged, almost ungaged and gaged conditions for the estimation of model parameters (*N* corresponds to the number of flow observations under which the application of classic estimation procedures are no longer robust)

After presenting the methodology (Section 2), we present the data set of 609 catchments and the two hydrological models used to evaluate the proposed approach (Section 3). The results are then presented and discussed in Section 4 and conclusions and perspectives are discussed in Section 5.

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

2 Presentation of the parameter estimation method for almost ungaged catchments

This section outlines the proposed approach and describes how it blends regional knowledge with point flow measurements in order to address the question of parameter estimation in almost ungaged catchments.

2.1 Origin of the method

The proposed method builds on two existing approaches that we have merged:

1. The neighborhood approach [see e.g. *Oudin et al.*, 2008] used for entirely ungaged basins, based on regional information: the parameter set for the target ungaged catchment is chosen among existing parameter sets, previously calibrated on gaged catchments. A distance between the ungaged catchment and its gaged neighbors can be defined either geographically (spatial proximity) or in the space of catchment descriptors (physical similarity).
2. The DIScrete Parameterization (DISP) approach [*Perrin et al.*, 2008] proposed for gaged catchments: as in the neighborhood approach, the parameter set for the ungaged catchment is selected within a collection of existing parameter sets, but here the criterion for parameter set selection is the value of the model error obtained by running the model with an existing parameter set on the target catchment. This method has been preferred over classic optimization algorithms since *Perrin et al.* [2008] showed that they lead to overfitting situations on short calibration time series.

These two approaches use the same prior information (a library of parameter sets previously calibrated on gaged catchments) but differ in the way parameter sets are selected, i.e. the way they define the distance between the donor gaged catchment and the target ungaged catchment: in the first case, the distance is either defined in the space of physical descriptors or geographically (regional information); in the second case, the distance is a function of the difference between model simulations and flow observations (local information).

2.2 Prerequisites

In the presentation below, we consider that the lumped continuous rainfall-runoff model we wish to apply has previously been calibrated on *p* gaged catchments, providing *p* parameter sets that constitute a parameter library. For these *p* catchments and for the ungaged catchment studied, a number of physical descriptors (e.g. catchment area, drainage density, mean slope, vegetation cover, etc.) and geographic coordinates are also available. Based on this regional information, the proposed approach can be applied using either physical similarity or spatial proximity as a basis for hypothesizing catchment similarity.

In these neighborhood methods, a pool of *M* best parameter sets instead of a single best set can be used, since this often gives better model results [*Goswami et al.*, 2007]. Pools of parameter sets are generally applied by averaging the corresponding simulated model outputs: the model is applied on the ungaged catchment with each of the *M* parameter sets before averaging the *M* outputs. This option indeed performs better than averaging the *M* parameter sets before applying the model, given the non-linearity of simulated processes [*Oudin et al.*, 2008].

In terms of data availability, we assume that the continuous time series of areal precipitation and potential evapotranspiration (PE) are available in the ungaged target catchment over the period where individual flow measurements are made, allowing the continuous rainfall-runoff model to be run over this period.

2.3 Proposed approach

The method outlined below was found to be the most efficient among many variants that are not presented here for the sake of brevity. *Rojas-Serna* [2005] provides full details of these other attempts, e.g. the design of weighted objective functions based on model errors and the departure of parameter values from their initial estimates obtained by regionalization. Note that the framework proposed below is presented using flow as a source of information but could also be generalized to other variables deemed useful for parameter estimation (soil moisture, snow cover, etc.). The proposed approach involves three steps:

Step 1: Ranking the library's *p* parameter sets

The parameter sets of the *p* gaged catchments in the library are first ranked using the selected neighborhood approach. The closest catchment is given rank 1, the most remote is given rank *p*. We note r_j^{reg} the rank of the j^{th} parameter set in the library. It is up to the end-user to define the neighborhood approach and the associated distance metric.

In parallel, the model is run on the ungaged catchment using each parameter set of the library in turn and the available precipitation and PE time series. For each flow simulation produced, a model error *F* is calculated on the *N* dates when flow observations are available. It is up to the user to define the formulation of *F*, given his knowledge of the model and his modeling objectives. The *p* parameter sets are ranked by increasing model error. The parameter set providing the lowest model error is given rank 1, the one providing the largest error is given rank *p*. We note r_j^{loc} the rank of the j^{th} parameter set in the library.

Step 2: Combining ranks and selecting a pool of *M* parameter sets

Two distances between the target catchment and the catchments in the library were defined in the previous step: a distance in terms of neighborhood and a distance in terms of model error. Both distances contain information on the relevance of the donor catchments and we wish to combine them into a single index. As the distances are not expressed in the same unit and do not vary over the same range in both cases, it is difficult to use their absolute values in the combination. After testing various solutions [*Rojas-Serna*, 2005], the option to combine the previously defined ranks was found to be a good compromise between simplicity and efficiency. A linear combination of ranks is made:

$$r_j = \alpha . r_j^{reg} + (1-\alpha) . r_j^{loc} \qquad \text{Eq. (1)}$$

where r_j is the combined rank of the j^{th} parameter set in the library and α is a weighting coefficient (varying between 0 and 1) expressing the relative importance of the regional information compared to the local information. Its value needs to be determined empirically (see section 4.3). When α equals 1, the method comes down to the neighborhood approach. When α equals 0, the method only uses point flow information. The combined rank r takes values between 1 and p. Note that two parameter sets in the library may have the same combined rank, which means that they will be considered equivalent in the proposed method.

This procedure merges the regional information with the local information gained from point flow measurements and guides the selection of donor catchments that are eventually both regionally and locally relevant. Using the combined rank r, a pool of the M closest parameter sets can be selected (i.e. the M parameter sets ranked first). The choice of M will be discussed in Section 4.2.

Step 3: Determining the flow in the almost ungaged catchment using the pool of *M* parameter sets

Using the available precipitation and PE time series of the almost ungaged catchment, M flow time series are simulated by the model on the almost ungaged catchment using each M

selected parameter set. Then the M simulated series are averaged (output averaging) to obtain the flow simulation for the target catchment.

2.4 Setting rules for applying the method

To practically apply the method when N flow measurements are available, two values must be defined:

1. the number M of parameter sets selected in the pool to be applied on the target catchment;
2. the value of α, i.e. the weighting coefficient that defines the relative importance of regional information.

We can expect the values of α to depend on the number N of flow measurements. Indeed, for large values of N, there will be a lot of information in the flow data, so that we ought to give more weight to the minimization of model error than to the neighborhood. Consequently, α should take values closer to 0. The sensitivity of the proposed approach to M and α values is analyzed in Section 4.

2.5 Options for case study application

In the case study application described in Section 3, the following choices were made. We used spatial proximity as the neighborhood approach, since it is the best performing method on our data set [*Oudin et al.*, 2008]. To compute the distance separating the neighboring catchment from the target ungaged catchment, a distance combining the horizontal Euclidean distance between the outlets (d_{outlet}) and the horizontal Euclidean distance between the centroids ($d_{centroid}$) was selected. Previous tests [*Lebecherel*, 2015] showed that using this distance was beneficial to transfer information between catchments of different sizes. The distance d considered is therefore defined as:

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

$$d = 0.2 \times d_{outlet} + 0.8 \times d_{centroid} \qquad \text{Eq. (2)}$$

In terms of model error *F*, here we selected the root mean square error (RMSE) calculated on all the available flow measurements:

$$F = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\sqrt{Q_i} - \sqrt{C_i}\right)^2} \qquad \text{Eq. (3)}$$

where Q_i and C_i are the observed discharge and the discharge calculated with the model, respectively, for the date of the i^{th} flow measurement. The root square transformed flows were used to compute *F*, because *Oudin et al.* [2006] showed that this formulation limits the influence of high flows and provides a more general model.

Obviously, the method can be applied with other options according to the modeler's choice, in terms of neighborhood, distance calculation [*Gottschalk et al.*, 2011] or model error formulation [*Crochemore et al.*, 2015].

3 Data, models and assessment methodology

For a general evaluation of the method, we used a large set of catchments and two rainfall-runoff models.

3.1 Catchments and data set

The proposed approach was tested on a large and varied catchment set for two reasons. First, when a method is assessed on a large data set, we can have greater confidence in its applicability and robustness [*Andréassian et al.*, 2006; *Gupta et al.*, 2014]. Second, the method itself requires a library of parameter sets that should be large enough to be representative of the conditions of the zone studied [see the discussion by *Perrin et al.*, 2008].

We used a set of 609 small to medium-size catchments in France (Figure 2). This set represents a variety of hydrometeorological conditions, as shown in Table 1. Catchments were selected to have limited gaps in flow series (less than 10% for every year of the series) and limited snow influence. For the sake of brevity, physical catchment descriptors are not detailed here since they were not used in the tests, but the catchment set includes various physical conditions.

Quantiles	0.05	0.25	0.5	0.75	0.95
Catchment area (km²)	34	109	270	833	4514
Mean elevation catchment (m)	87	180	375	781	1350
Annual rainfall, P (mm/yr)	714	863	1003	1209	1688
Annual potential evapotranspiration (PE) (mm/yr)	533	616	655	687	782
Annual discharge, Q (mm/yr)	159	272	411	643	1308

Table 1. Main characteristics of the 609 catchments used to test the approach.

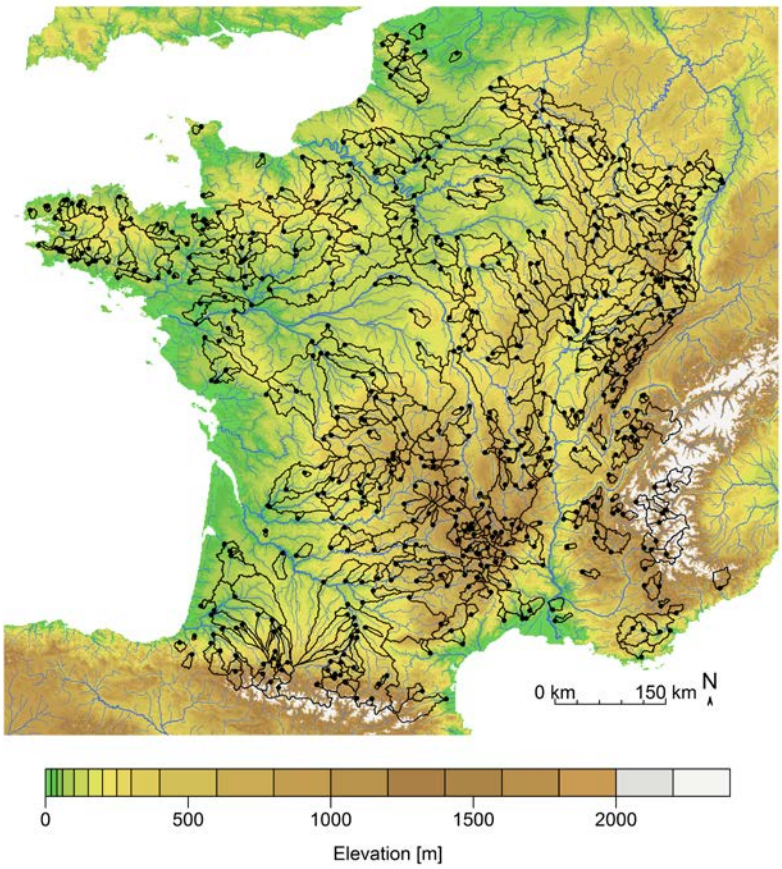

Figure 2. Location of the 609 French catchments used in this study (dots indicate the gauging stations and solid lines the catchment boundaries)

Daily data were available over the 1996–2005 period. Areal catchment rainfall was calculated using the SAFRAN gridded values provided by Météo-France [*Vidal et al.*, 2010]. Mean annual curves of potential evapotranspiration were computed using the formula provided by *Oudin et al.* [2005] based on air temperature and extra-terrestrial radiation. Streamflow time series were extracted from the HYDRO national archive (http://hydro.eaufrance.fr). These data are generally considered to be of good quality.

3.2 Rainfall-runoff models and parameter library

The method outlined in this paper can be applied with any lumped rainfall-runoff model. Here we used two models to reach more robust conclusions and possibly analyse differences between models. We applied the four-parameter GR4J model [*Perrin et al.*, 2003] and the eight-parameter TOPM model [*Michel et al.*, 2003]. A sketch of the model structures is shown in Figure 3 and the meaning of the parameters is given in Table 2. These two models were previously used in French catchments showing comparable levels of performance [*Oudin et al.*, 2008; *Perrin et al.*, 2008].

	Parameter	**Meaning**
GR4J	X1	Capacity of the production store (mm; positive)
	X2	Water exchange coefficient (mm; positive or negative)
	X3	Capacity of the nonlinear routing store (mm; positive)
	X4	Unit hydrograph time base (day; ≥ 0.5)
TOPM	X1	Parameter of the exponential store (mm; positive)
	X2	Evapotranspiration parameter (mm; positive or negative)
	X3	Topographic index distribution parameter (mm; positive)
	X4	Pure time delay (day; ≥ 1)
	X5	Topographic index distribution parameter (mm; positive or negative)
	X6	Evapotranspiration parameter (mm; positive)
	X7	Capacity of the interception store (mm; positive)
	X8	Capacity of the routing store (mm; positive)

Table 2. List of parameters of the GR4J and TOPM models

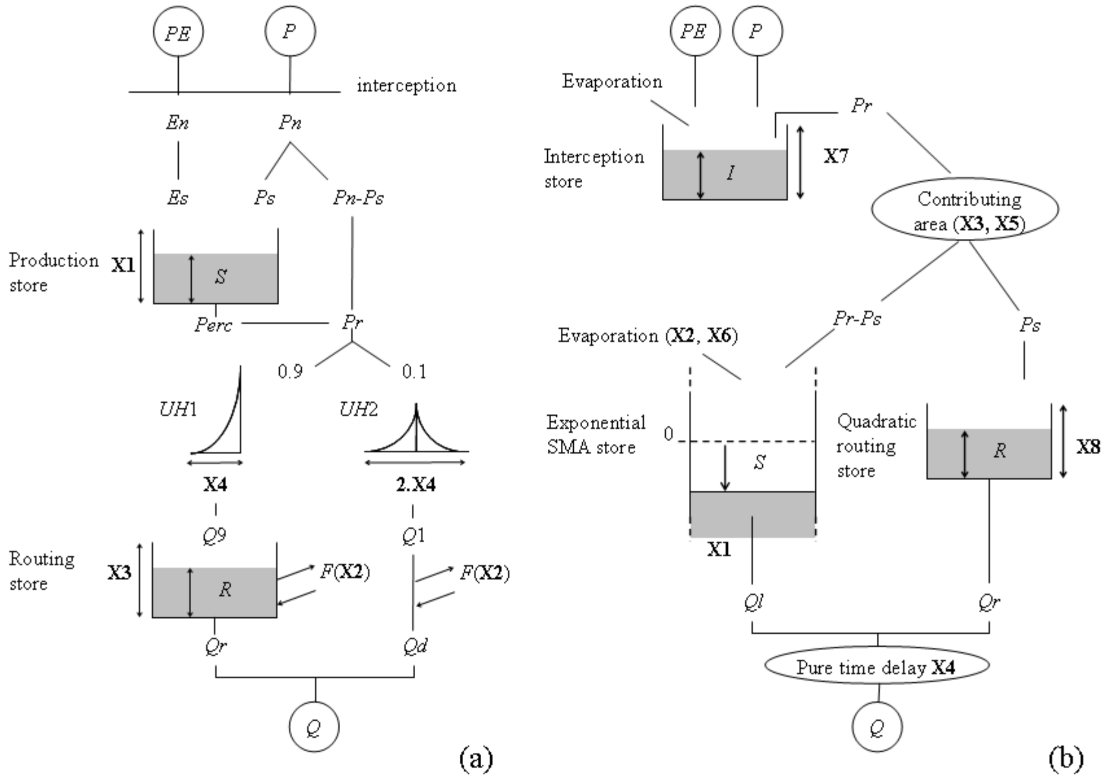

Figure 3. Schematic diagrams of the GR4J (a) and TOPM (b) rainfall-runoff models (PE: potential evapotranspiration; P: precipitation; Q: streamflow; Xi: model parameter i; other letters are internal state variables)

To build the library of model parameters, the parameters were calibrated on each catchment using the optimization algorithm applied by *Edijatno et al.* [1999]. The objective function used during optimization is the *Nash and Sutcliffe* [1970] criterion calculated on root square transformed flows, consistently with the formulation of *F* chosen here (see Section 2.5). Other objective functions [e.g. *Gupta et al.*, 2009] could be used consistently with the choices made in applying the method.

3.3 Assessment procedure

The proposed approach was successively applied to each catchment considered in turn as ungaged. Each time, the parameter set of the catchment under study was excluded from the library to test the approach.

We applied the split-sample test scheme advised by *Klemeš* [1986] by splitting the available record into two periods (1995–2000 and 2000–2005) that were alternatively used for model parameter identification and model assessment in validation (parameter identification in period 1 and validation in period 2 and vice-versa). For each period, the first year (1995 and 2000, respectively) was used for model warm-up, which means that model performance was actually computed in the 1996–2000 and 2001–2005 periods.

Here the flow measurements were randomly drawn in the flow series on each period. The random option was chosen because it corresponds quite well to the case where one collects point flow data without following a predefined acquisition strategy. This can be considered as the "poor-man's" option, i.e. a baseline strategy. The flow data were drawn incrementally: a new flow measurement drawn in the flow series is added to the set of already selected flows, mimicking what happens in operational conditions (i.e. a sample of k+1 measurements includes the sample of k measurements already made). Hence it is considered that the flow information is consistently increasing when N increases. The random selection was made once per catchment. Since the number of catchments is large, this does not prevent obtaining robust results.

More advanced sampling strategies could be adopted to improve modeling efficiency [*Viviroli and Seibert*, 2015], but this was not within the scope of this article.

3.4 Evaluation criteria

The evaluation of the method was based on model performance obtained in validation mode as measured by the *Nash and Sutcliffe* [1970] criterion (*NS*). *NS* varies within the interval]-∞, 1]. The lack of a lower bound for this criterion is a problem when working on a large set of catchments in ungaged conditions, because the criterion may take highly negative values in some catchments where the model fails. This prevents making meaningful averages of efficiency criteria over the catchment set. To circumvent this pitfall, we used the *C* criterion proposed by *Mathevet et al.* [2006] and given by:

$$C = \frac{NS}{2 - NS} \qquad \text{Eq. (3)}$$

This transformation is bounded in the interval]−1, 1], which allows making meaningful performance averages over the test set. Note that this criterion keeps the same baseline as the Nash-Sutcliffe criterion ($C = 0$ when $NS = 0$), has the same optimum (1 means perfect simulation for both criteria), but yields lower positive values compared to the *NS* criterion (e.g. $C = 0.67$ when $NS = 0.80$).

The efficiency of the parameter estimation method will be assessed by mean *C* values obtained by the hydrological model over all validation tests (here 1218, i.e. twice the number of catchments). *C* was calculated on root square transformed flows (C_{RQ}), like the objective function, but also on flows (C_Q) and logarithm-transformed flows (C_{LQ}), to put more weight on high and low flows than C_{RQ}, respectively [see e.g. *Pushpalatha et al.*, 2012].

3.5 Reference methods

The model performance obtained by applying the proposed method was compared with the model performance obtained:

1. in fully ungaged conditions: we applied the approach of spatial proximity with the output averaging the pooling option [*Oudin et al.*, 2008];
2. in fully gaged conditions: we applied the optimization algorithm mentioned above using all the data available in the calibration period.

Other benchmarks could be considered in model evaluation [see e.g. *Seibert*, 1999], but since the intention here was not to compare various parameter estimation methods, we kept only these two “extreme” benchmarks.

4 Results

In this section, we present the results of the proposed approach for almost ungaged catchments. We discuss the sensitivity of the method to the value of the weighting factor of the regional information (α) and to the number (M) of parameter sets in the pool applied to the ungaged catchment. The two hydrological models were applied with an increasing number of flow measurements (N). After determining the optimum values, we evaluate the performance of the proposed approach, which we compare with the reference methods.

4.1 Results of reference methods

For the two models (GR4J and TOPM) and three evaluation criteria (C_{LQ}, C_{RQ} and C_Q), Figure 4 shows the performance distributions obtained on the catchment set in calibration and validation using the full flow data (fully gaged conditions) and when applying the neighborhood approach (fully ungaged conditions) for an increasing number of donor catchments. The best results in ungaged conditions are obtained with a small pool of parameter sets: seven parameter sets for GR4J and nine parameter sets for TOPM. This difference between the two models is consistent with the results found by *Oudin et al.* [2008] on a similar data set. The smoothing effect of the output averaging option probably explains why no significant improvement is found with more donors on average.

The efficiencies obtained by the two models are close in ungaged conditions. GR4J is slightly better than TOPM in fully gaged conditions. Note that in gaged mode, the models are quite efficient on average on the 609 catchments, since they reach C efficiencies of around 0.60–0.65 (equivalent to 0.75–0.79 in terms of the NS criterion). As expected, the efficiency obtained in fully gaged conditions is far better than that obtained in fully ungaged conditions. This sets the range of improvements that can be obtained using flow information.

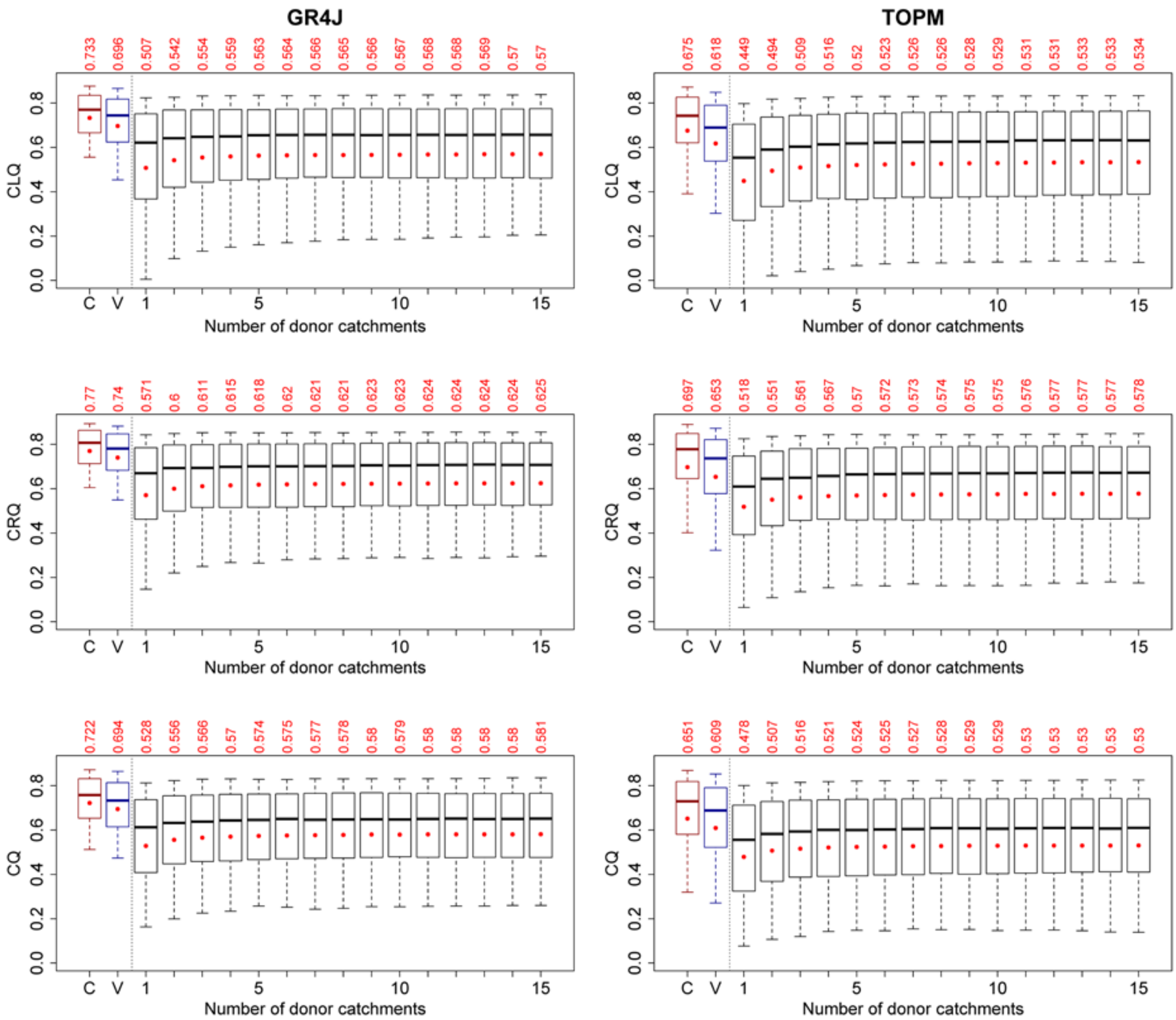

Figure 4. Distribution of model efficiency in calibration and validation in gaged conditions (red and blue boxplots, respectively) and in validation in ungaged conditions with the number of donor catchments (boxplots show the 0.1, 0.25, 0.50, 0.75 and 0.9 percentiles; the mean value is indicated on top of the frame and by the cross).

4.2 Sensitivity to the number M of neighbors in almost ungaged conditions

We evaluated the sensitivity of the proposed approach to the number M of parameter sets selected in the pool applied to the ungaged catchment. To simplify the presentation, we fixed the weight of regional information, α, at 0.5 (the results are very similar for other α values). We tested the two models using an increasing number of donors (M = 1, 2, …, 9, 10, 15) and an increasing number of flow measurements (N = 1, 2, …, 6, 7, 10, 20, 50).

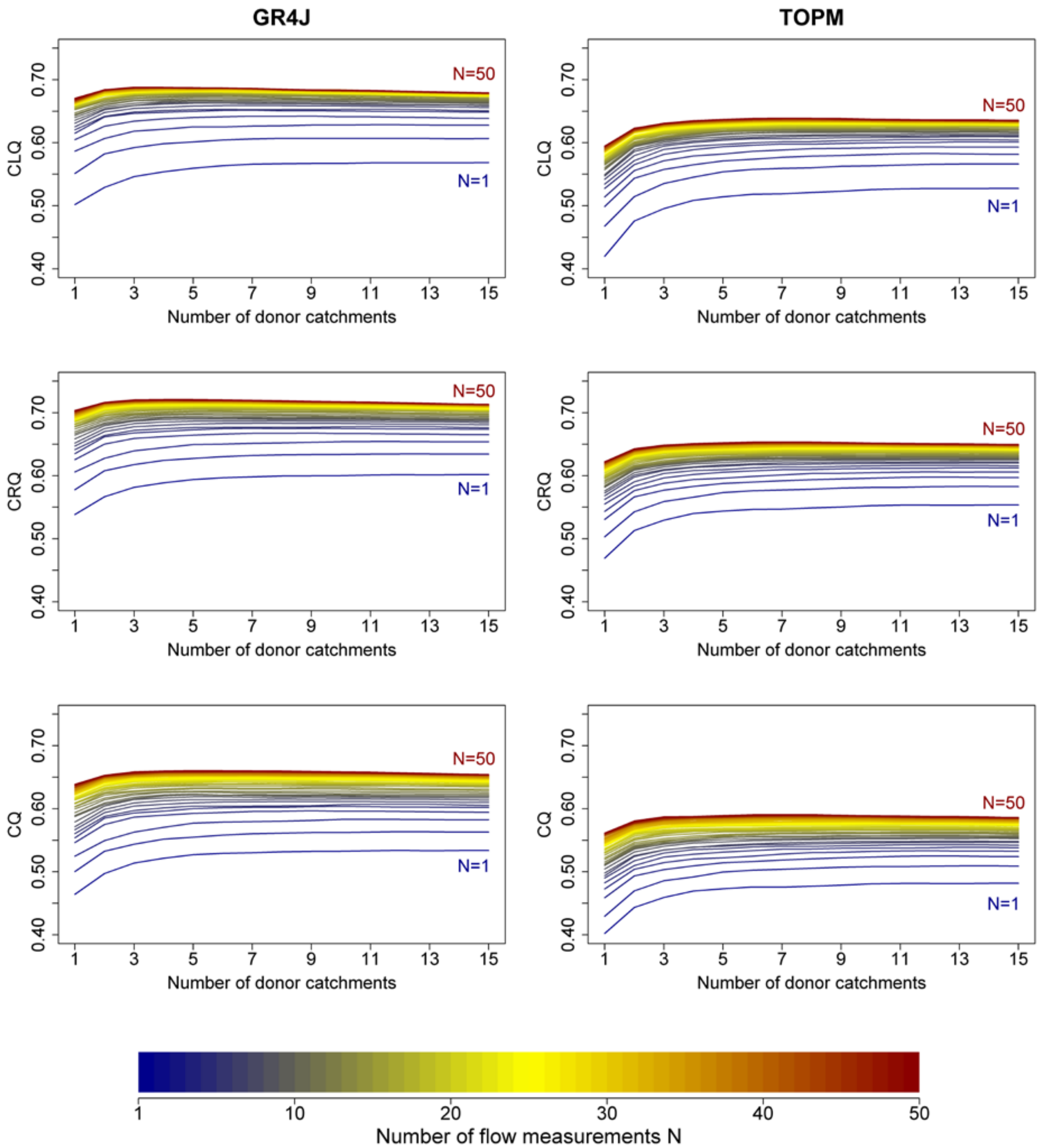

Figure 5. Mean model efficiency (C_{LQ}, C_{RQ} and C_Q) in validation over the catchments with the number *M* of donor sets in almost ungaged conditions for the GR4J and TOPM models. Each line represents the application of the method with a given number of flow measurements *N* (from 1 to 50). Regional and local parameters are equally weighted ($\alpha = 0.5$ in Eq. 1).

Figure 5 presents the mean model efficiency (C_Q, C_{LQ} and C_{RQ}) over the catchments with the number *M* of donor sets in almost ungaged conditions for the GR4J and TOPM models for an increasing number of flow measurements used for parameter estimation. It shows that as soon as a few donor sets are used, the results stabilize and are no longer sensitive to *M*. The same behavior is observed for any number N of point flow measurements, the model or the criterion. For a given model, the optimum number of donors does not seem to depend much on N. Therefore we chose to set the number of donor sets to the same value identified in ungaged conditions, i.e. M = 7 for GR4J and M = 9 for TOPM. These settings were used in subsequent tests.

4.3 Sensitivity to the weighting factor α

We evaluated the sensitivity of the proposed approach to the value of weighting factor, α, which sets the weight of the regional information. We can expect that the larger the number of point flow measurements, the lower the value of α, given that the flow measurements will yield more information. We tested values of α from 0 (no use of regional information) to 1 (no use of local information) (see Eq. 1). α was varied between 0 and 1 by steps of 0.1 and each value of α was tested using different numbers of flow measurements (from 1 to 50, as in the previous section).

Figure 6 presents the mean model efficiency (C_Q, C_{LQ} and C_{RQ}) over the catchments with the weighting factor α for an increasing number of flow measurements (from 1 to 50). When α equals 1 (i.e. using only regional information), the results are the same as those found previously for the strictly ungaged catchments. When α equals 0 (i.e. only considering local flow information), model efficiency progressively increases when N increases. The performance exceeds the performance obtained in ungaged conditions as soon as a few (actually five at most) flow measurements are available for GR4J and TOPM.

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

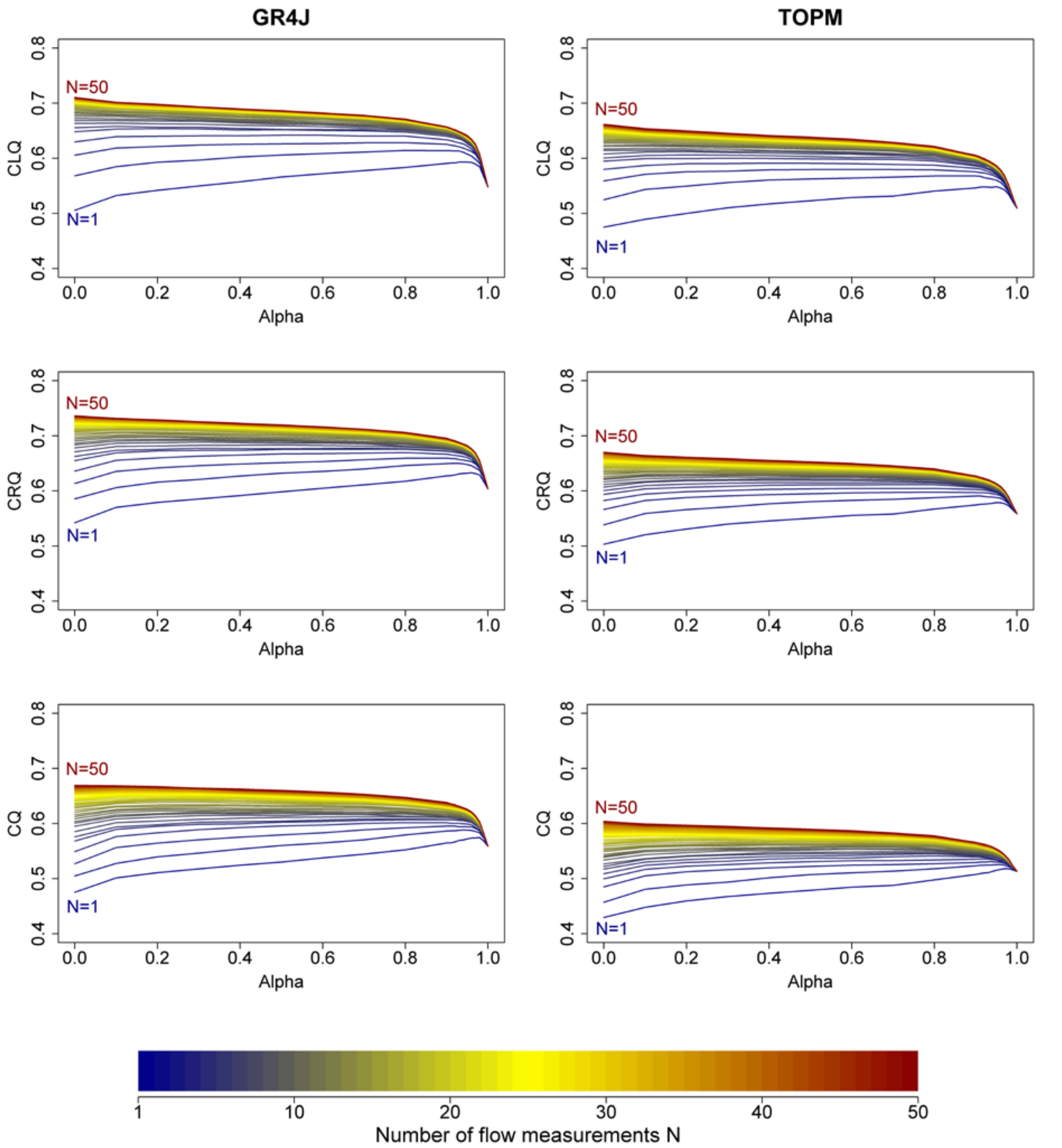

Figure 6. Mean model efficiency (C_{LQ}, C_{RQ} and C_{Q}) in validation over the catchments with the weighting factor α of the regional information (α=0: no regional information; α=1: only regional information) in almost ungaged conditions for the GR4J and TOPM models. Each line represents the application of the method with a given number of flow measurements *N* (from 1 to 50). The number *M* of donor catchments is set to seven and nine for GR4J and TOPM, respectively.

Figure 6 also shows that an optimum value of α can be identified in the efficiency curves drawn for each value of *N*. This value is very close to 1 when a single flow measurement is available, which means that it is better to mostly trust regional information in these conditions. However, it does not strictly equal 1: this indicates that some information may be gained even from a single flow measurement. Note that for all values of N, the increase in model performance when α departs from 1 is very rapid. This means that flow measurements yield very different and complementary information compared to solely regional information (i.e. that the rankings are quite different). This can be linked to the finding of *Vrugt et al.* [2002] and *Seibert and Beven* [2009], who found that even a few flow measurements already contain valuable information.

When N increases, the optimum α value progressively decreases. This can be expected, since the regional information becomes less and less relevant in comparison to the local (flow) measurements. For each value of N, we identified this optimum value (based on the mean results over the catchment set) and plotted it against N (see Figure 7). The patterns are quite similar for both models, with slight differences, which may come from the fact that the neighborhood approach does not give equally valuable information in each case. The differences between criteria are greater than between models, but consistent between the two models. When N increases, the weight of regional information drops faster for C_{LQ} than C_Q (with intermediate results for C_{RQ}): approximately ten measurements are necessary to get α close to 0 for C_{LQ}, while at least 30 values are necessary for C_Q. This may indicate that regional information is worse and/or that fewer flow measurements are necessary when focusing on low-flow conditions.

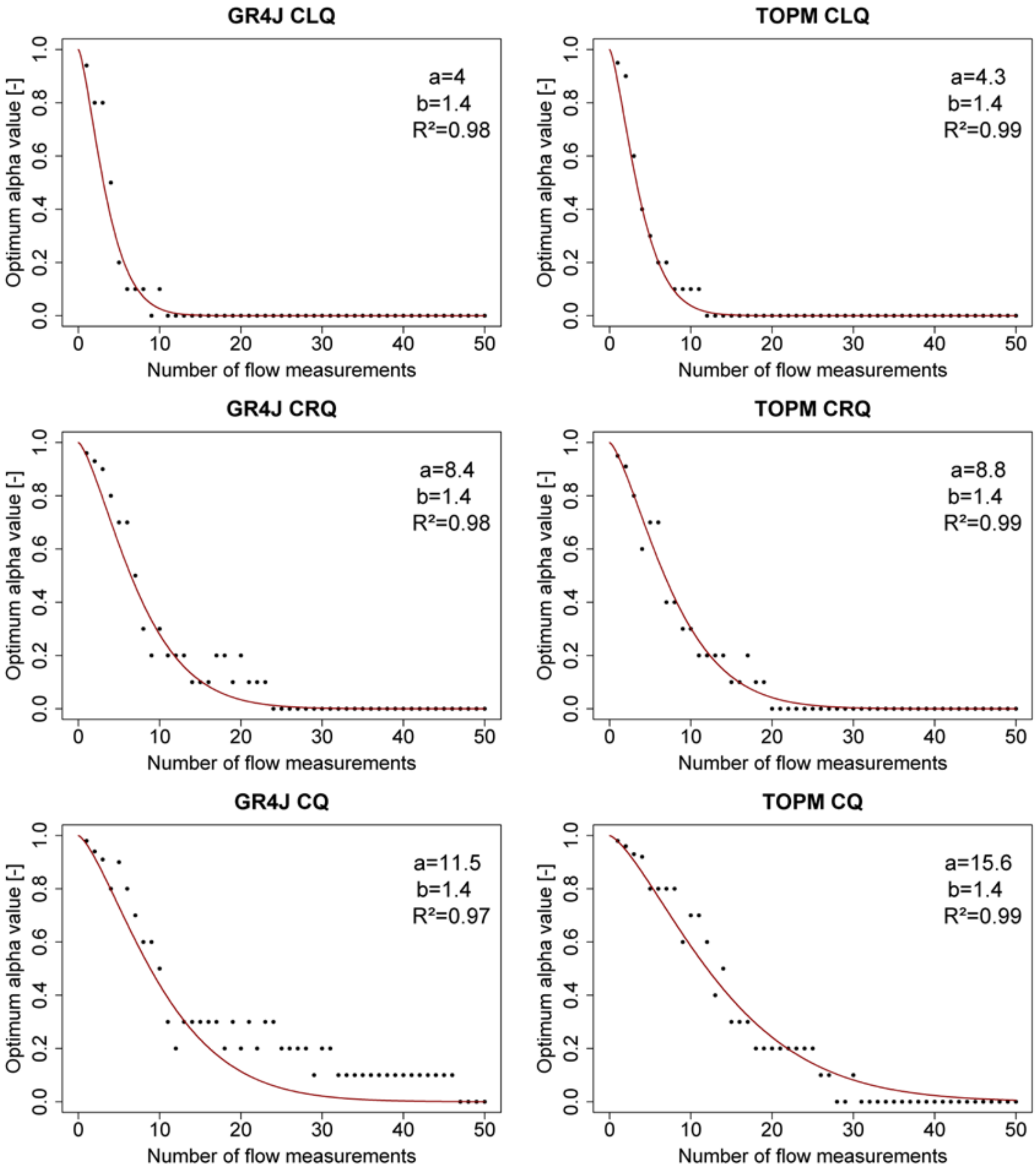

Figure 7. Optimum values of the weighting factor α of the regional information in almost ungaged conditions for the GR4J and TOPM models for increasing numbers of flow measurements (from 1 to 50). The number N of donor catchments is set to seven and nine for GR4J and TOPM, respectively. The solid line corresponds to the curve defined in Eq. (4). α values were discretely tested with a 0.1 step.

It therefore seems that the optimal settings of the proposed approach in terms of the relative importance of the regional information may partly depend on the model and, to a greater extent, on the modeling objectives. However, given that the shapes of the relationship between α and N are similar, a general formulation to determine the value of α for a given number N of flow measurements available is proposed:

$$\alpha = \exp\left[-\left(\frac{N}{a}\right)^{b}\right] \qquad \text{Eq. (4)}$$

where a and b are two coefficients to be determined for the model and objectives selected. The values of a and b and the corresponding curves are shown in Figure 7. Additional tests (not detailed here) show that b is the less sensitive parameter of the two and the results indicate that fixing it does not yield significant loss in modeling efficiency. In the subsequent tests, we used the value b=1.4, which was found to provide the best results.

4.4 Performance of the proposed approach

Figure 8 shows the changes in the mean efficiency of the GR4J and TOPM models obtained by applying the proposed approach for almost ungaged conditions to our data set, using between 1 and 50 point measurements. It exploits both local flow measurements and regional information, using the number M of neighbors determined previously and the optimum values of α calculated by Eq. 4. It is compared with the two reference methods (the neighborhood for the fully ungaged case and the optimization algorithm for the fully gaged case).

On average, the proposed approach efficiently uses the information provided by regional information and local flow measurements, since it proves more efficient than the reference method (i.e. the fully ungaged case) for all the efficiency criteria (C_Q, C_{LQ} and C_{RQ}).

These results show that the method proposed here can effectively exploit the two sources of information. It is particularly valuable when only a few flow measurements are available. Some improvement in model performance can be obtained in comparison with the sole use of regional information when only one or two flow measurements are available. In most cases (models and criteria), ten measurements (or a few more in the case of C_Q) can reduce the performance gap between the gaged and ungaged situations by more than 50%.

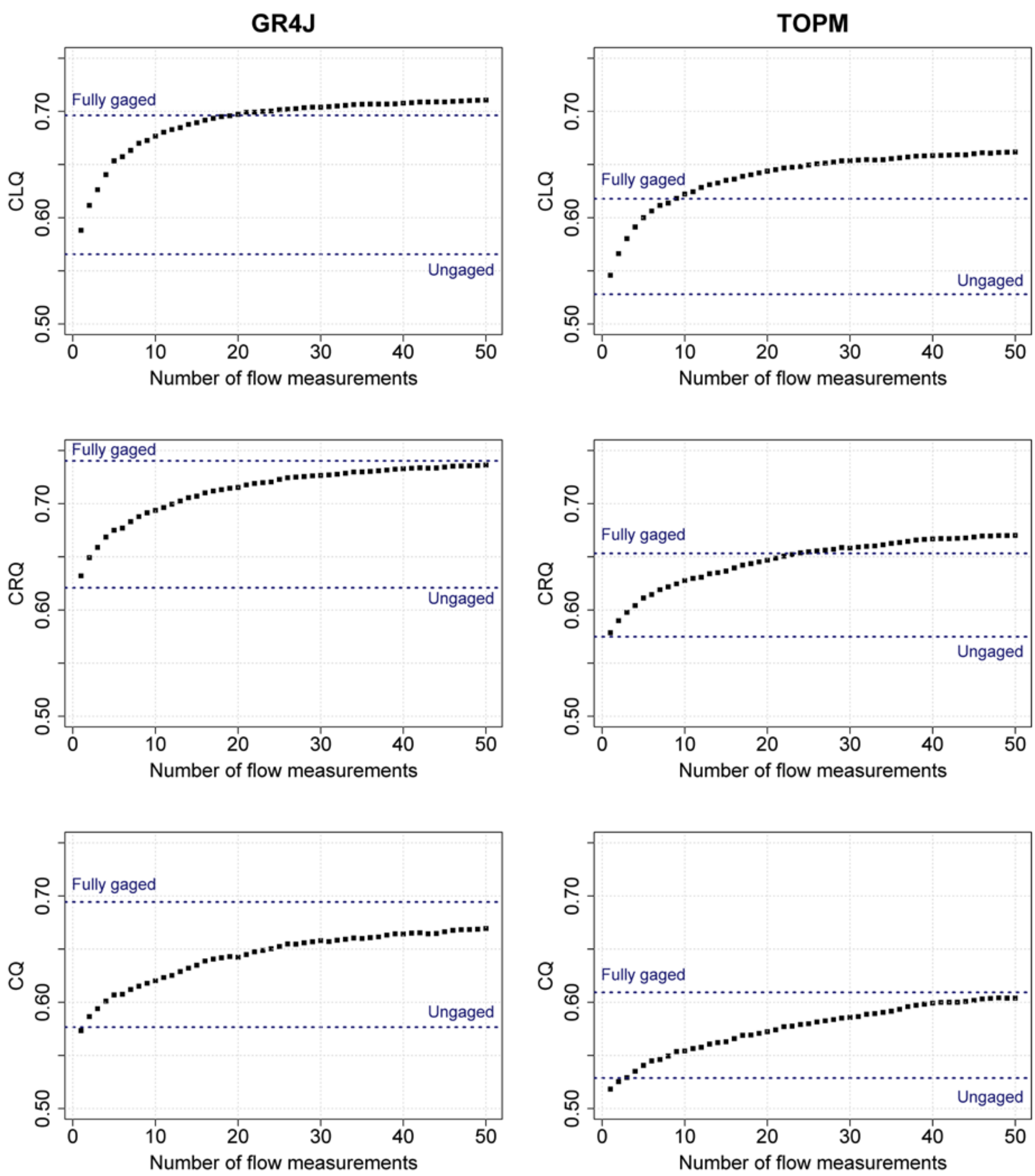

Figure 8. Efficiency of the proposed approach in validation for almost ungaged catchments compared to the ungaged and gaged conditions for the GR4J and TOPM models for an increasing number of flow measurements (from 1 to 50). The number N of donor catchments is set to seven and nine for GR4J and TOPM, respectively.

For the C_{LQ} criterion, the proposed approach provides more efficient results than the fully gaged case for both models when N exceeds 10–20. This may appear surprising, but can partly be explained by the fact that the evaluation criterion is different from the calibration objective function (C_{RQ}). Here, the fully gaged value is not the optimal validation value for this criterion. More surprisingly, a similar behavior can be observed for the TOPM model and the C_{RQ} evaluation criterion. These results may be explained by the use of multiple sets of parameters (a pool of M=9 for TOPM) to simulate flows on almost ungaged catchments. Figure 9 shows the efficiency of TOPM when applying the method for M=9 and M=1 (i.e. using a single donor). Clearly, the multi-parameter approach outperforms the single-parameter approach for TOPM, which remains below the fully gaged case. Note also that the results shown here were obtained in validation mode, and the multi-parameter approach may be more robust than the single-parameter approach. This behavior is observed to a larger extent for TOPM than for GR4J, which may come from its higher number of parameters and consequently lower robustness due to possible equifinality [as discussed by *Perrin et al.*, 2008].

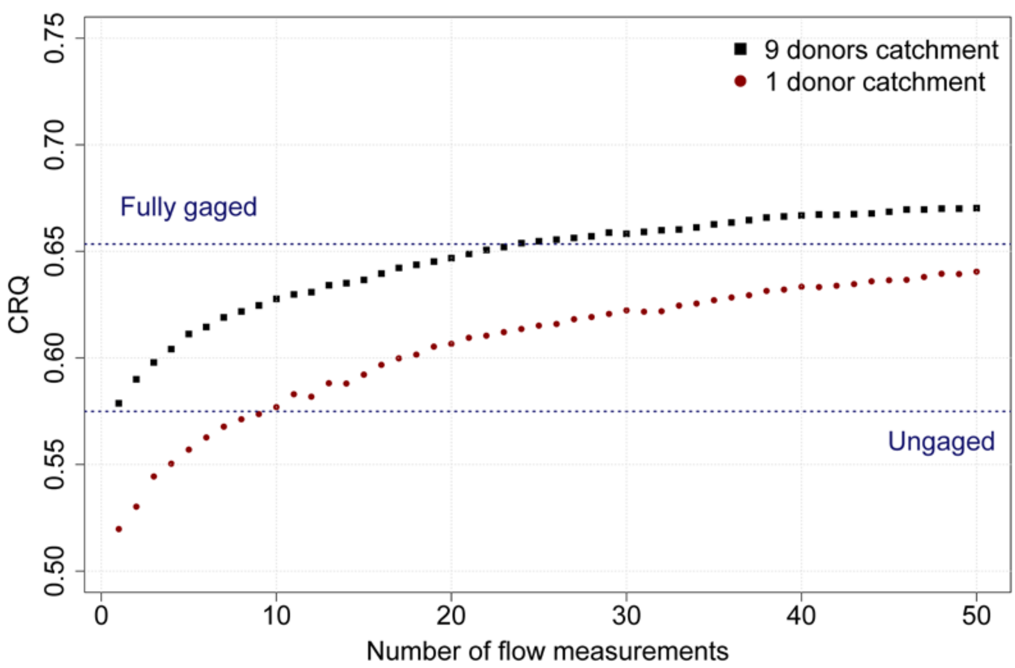

Figure 9. Efficiency of the proposed approach for the TOPM model and for the C_{RQ} evaluation criterion in validation, for increasing numbers of flow measurements (from 1 to 50) and two donor catchment values (M) (squares for nine donors and dots for one donor).

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

4.5 Impact on parameter estimates

We showed that progressively adding local information to existing regional information yields better model performance in validation. This means that estimated parameters are more transferable in time. Hence one could expect that they are closer to values estimated using full flow information, which should be more representative of the catchment behavior over the long term. Therefore, the proposed approach should help obtain more general parameters.

This was investigated by analyzing the parameter variability when the number of point flow measurements increases. Because parameters can vary over a wide range, parameter values were normalized in each case by the long-term optimum parameter set, i.e. the parameter set obtained by calibration on the full data period. The distribution of these normalized parameters on the sample of 609 catchments is shown in Figure 10 for *N* between 1 and 50. To simplify the analysis, only the parameter set ranked first in our approach was considered here (trends are similar when using the pool of parameter sets). For sake of brevity, we only show the results for the GR4J model. Although the trend is smooth, one can see that the variability decreases (the inter-quantile ranges decrease) for most parameters when *N* increases. This means that on average, the parameter values tend to come closer to the long-term optimum value. No real trend is observed for the X4 parameter (time base of the unit hydrograph), but this parameter shows the least variability, i.e. it is probably better determined by regionalization than the others. Note that median values of each boxplot are very close to 1, indicating that the procedure does not tend to introduce any significant bias in the estimated parameter values on average.

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

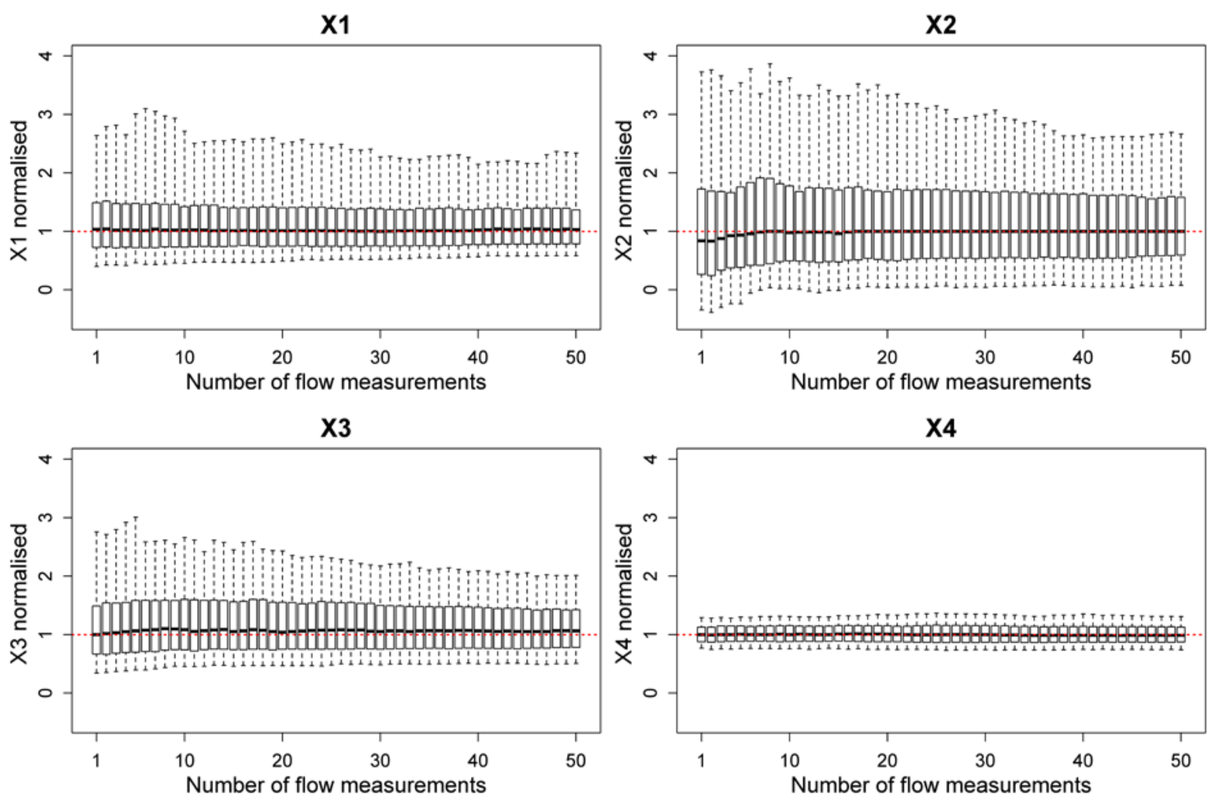

Figure 10. Distribution of normalized best parameters for the GR4J model over the catchment set with the increase of flow measurements (the meaning of parameters is given in Table 2). Boxplots show the 0.1, 0.25, 0.50, 0.75 and 0.9 percentiles.

5 Conclusion and perspectives

In this article, a combined approach was devised to simultaneously exploit regional information and local measurements. The proposed approach weights the two sources of information depending on the availability of point flow measurements. It provides simulations intermediate between fully ungaged and fully gaged situations. In this sense, the method intends to make the connection between these two cases. This approach opens ways to make model parameter estimation more reliable in all the catchments where only a few flow measurements are available. As more flow measurements become available, the model becomes more robust, i.e. it obtains better results in validation. The parameters also tend to progressively converge to the long-term values estimated in fully gaged conditions. Hence the procedure progressively adapts the values of parameters as new flow information is added, in some type of "learning process" [*Buytaert and Beven,* 2009].

Starting from a prior (the regional estimate) and progressively narrowing the parameter values by adding new information may be seen as a typical Bayesian process [*Thiemann et al.*, 2001]. The advantage of the proposed approach is that it remains very simple without making statistical assumptions, but it would be interesting to compare it with formal Bayesian approaches in future work.

The method was tested using one specific regionalization approach, two rainfall-runoff models and three evaluation criteria targeting various flow types (high or low). Interestingly, the method’s settings were shown to depend more on the modeling objective than on the model used. However, the method is general and flexible enough to be efficiently applied to other models and objectives and a simple way was proposed to weight the regional and local information. The approach was tested on a large data set of 609 catchments and compared to two simple benchmarks, which gives confidence on its generality, overall efficiency [*Andréassian et al.*, 2007] and applicability at the regional level [*Drogue and Plasse*, 2014].

One major conclusion is that acquiring flow information by point measurements and efficiently combining it with prior regional information can be very effective in improving parameter estimation, as already suggested by other authors [*Seibert and Beven*, 2009]. Studies comparing various parameter estimation strategies (and possibly involving other types of observations than flow) would be very useful to conduct in the context of almost ungaged catchments.

One limit of these tests is that flows were sequentially drawn at random in the existing series, irrespective of the flow magnitude, dynamics (rising limb or recession) or season. These various conditions are known to provide different informative content for the estimation of model parameters [*Wagener et al.*, 2003]. Several authors attempted to define strategies for acquiring point flow data in a modeling perspective [see e.g. *Clausen*, 1995; *Juston et al.*, 2009; *Konz and Seibert*, 2010; *Singh and Bardossy*, 2012; *Viviroli and Seibert*, 2015]. This could also be investigated in the context of the approach proposed here. Ultimately, this could guide practicing hydrologists and operational gauging staff in defining gauging priorities and making decisions in real time.

Last, we mentioned in the introduction the PUB decade, which gave way to the Panta Rhei decade on “Change in hydrology and society” [*Montanari et al.*, 2013]. It is likely that the notion of an *almost ungaged catchment* could be revisited and extended in this context. Typically a sudden change on a catchment will strongly impact its behavior and make it suddenly ungaged in the sense that no or not enough data are available to properly model it. Hence the interactions between transposition in time and space could be further investigated in this context.

6 Acknowledgements

The authors wish to thank Météo-France and SCHAPI for providing meteorological and hydrological data sets, respectively. Conacyt (Mexico) and ONEMA (France) provided financial support for the PhD research of the first and second authors, respectively, and are gratefully acknowledged.

7 References

Anctil, F., C. Perrin, and V. Andréassian (2004), Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models, *Environmental Modelling & Software*, *19*(4), 357-368.

Andréassian, V., A. Hall, N. Chahinian, and J. Schaake (2006), Introduction and Synthesis: Why should hydrologists work on a large number of basin data sets?, paper presented at

Large sample basin experiments for hydrological model parameterization: Results of the Model Parameter Experiment - MOPEX, Paris, IAHS Publication n°307, 1-5.

Andréassian, V., J. Lerat, C. Loumagne, T. Mathevet, C. Michel, L. Oudin, and C. Perrin (2007), What is really undermining hydrologic science today?, *Hydrol. Processes*, *21*(20), 2819-2822.

Andréassian, V., N. Le Moine, C. Perrin, M.-H. Ramos, L. Oudin, T. Mathevet, J. Lerat, and L. Berthet (2012), All that glitters is not gold: the case of calibrating hydrological models, *Hydrol. Processes*, *26*(14), 2206-2210.

Andréassian, V., F. Bourgin, L. Oudin, T. Mathevet, C. Perrin, J. Lerat, L. Coron, and L. Berthet (2014), Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency, *Water Resour. Res.*, *50*(10), 8356 - 8366.

Bhatt, V. K., and A. K. Tiwari (2008), Estimation of peak streamflows through channel geometry, *Hydrol. Sci. J.-J. Sci. Hydrol.*, *53*(2), 401-408.

Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije (Eds.) (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, 465 pp., Cambridge University Press.

Brath, A., A. Montanari, and E. Toth (2004), Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model, *J. Hydrol.*, *291*(3-4), 232-253.

Buytaert, W., and K. Beven (2009), Regionalization as a learning process, *Water Resour. Res.*, *45*, W11419.

Castiglioni, S., L. Lombardi, E. Toth, A. Castellarin, and A. Montanari (2010), Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Adv. Water Resour.*, *33*(10), 1235-1242.

Catalogne, C., E. Sauquet, and M. Lang (2014), Using spot gauging data to estimate the annual minimum monthly flow with a return period of 5 years, *La Houille Blanche*(4), 78-87.

Chopart, S., and É. Sauquet (2008), Usage des jaugeages volants en régionalisation des débits d'étiage, *Revue des Sciences de l'Eau*, *21*(3), 267-281.

Clausen, B. (1995), Discharge Data-Collection and Analysis Strategies in Low-Flow Studies, *Nordic Hydrol.*, *26*(3), 191-204.

Corbonnois, J., D. François, I. Arts, M. Sary, E. Gille, P. Gamez, A. Devos, and J. P. Decloux (1999), Répartition des écoulements en basses eaux. Aspects méthodologiques. (Spatialization of low water flows. Methodology), paper presented at XXII IUGG General Assembly, 18-30 July 1999, Birmingham (UK), 253-260.

Crochemore, L., C. Perrin, V. Andréassian, U. Ehret, S. P. Seibert, S. Grimaldi, H. Gupta, and J. E. Paturel (2015), Comparing expert judgement and numerical criteria for hydrograph evaluation, *Hydrol. Sci. J.*, *60*(3), 402-423.

Decloux, J. P., and M. Sary (1991), Campagnes d'étiage. Objectifs, traitement et valorisation des données, *Mosella*, *XXIII*, 121-134.

Drogue, G. P., and J. Plasse (2014), How can a few streamflow measurements help to predict daily hydrographs at almost ungauged sites?, *Hydrol. Sci. J.*, *59*(12), 2126-2142.

Edijatno, N. O. Nascimento, X. Yang, Z. Makhlouf, and C. Michel (1999), GR3J: a daily watershed model with three free parameters, *Hydrol. Sci. J.*, *44*(2), 263-277, DOI: 210.1080/02626669909492221.

Eng, K., and P. C. D. Milly (2007), Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges, *Water Resour. Res.*, *43*(1).

Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008), Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, *44*(6), W06419.

Fernandez, W., R. M. Vogel, and A. Sankarasubramanian (2000), Regional calibration of a watershed model, *Hydrol. Sci. J.*, *45*(5), 689-707.

François, D., and M. Sary (1990), Etude méthodologique des basses eaux. estimation minimale des points de mesure, 14 pp, Cegum, Université de Metz.

François, D., and M. Sary (1994), Etude méthodologique des débits d'étiage, 30 pp, Cegum, Université de Metz.

Gill, M. K., T. Asefa, Y. Kaheil, and M. McKee (2007), Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique, *Water Resour. Res.*, *43*(7), W07416.

Goodwin, T., and A. Young (2007), Reducing uncertainty for the estimation of hydropower potential - the use of local data, paper presented at British Hydropower Association Conference, 11 October 2007, www.british-hydro.org, Birnam, Australia.

Goswami, M., K. M. O'Connor, and K. P. Bhattarai (2007), Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment, *J. Hydrol.*, *333*(2-4), 517-531.

Gottschalk, L., E. Leblois, and J. O. Skoien (2011), Distance measures for hydrological data having a support, *J. Hydrol.*, *402*(3-4), 415-421.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1-2), 80-91.

Gupta, H. V., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian (2014), Large-sample hydrology: a need to balance depth with breadth, *Hydrol. Earth Syst. Sci.*, *18*(2), 463-477.

Hrachowitz, M., et al. (2013), A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrol. Sci. J.*, *58*(6), 1198-1255.

Hughes, D. A., M. Gush, J. Tanner, and P. Dye (2014), Using targeted short-term field investigations to calibrate and evaluate the structure of a hydrological model, *Hydrol. Processes*, *28*(5), 2794-2809.

Hundecha, Y., T. Ouarda, and A. Bardossy (2008), Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the "spatial" structures of the parameters within a canonical physiographic-climatic space, *Water Resour. Res.*, *44*(1), W01427.

Juston, J., J. Seibert, and P. O. Johansson (2009), Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Processes*, *23*(21), 3093-3109.

Kim, U., and J. J. Kaluarachchi (2009), Hydrologic model calibration using discontinuous data: an example from the upper Blue Nile River Basin of Ethiopia, *Hydrol. Processes*, *23*(26), 3705-3717.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*(1), 13-24.

Konz, M., and J. Seibert (2010), On the value of glacier mass balances for hydrological model calibration, *J. Hydrol.*, *385*, 238-246.

Konz, M., S. Uhlenbrook, L. Braun, A. Shrestha, and S. Demuth (2007), Implementation of a process-based catchment model in a poorly gauged, highly glacierized Himalayan headwater, *Hydrology and Earth System Sciences*, *11*(4), 1323-1339.

Koren, V., M. Smith, and Q. Duan (2003), Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, *Calibration of Watershed Models, Water Science and Application*, *6*, 239-254.

Kuzmin, V., D. J. Seo, and V. Koren (2008), Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search, *J. Hydrol.*, *353*(1-2), 109-128.

Laaha, G., and G. Blöschl (2005), Low flow estimates from short stream flow records - a comparison of methods, *J. Hydrol.*, *306*(1-4), 264-286.

Lebecherel, L. (2015), Sensibilité des calculs hydrologiques à la densité des réseaux de mesure hydrométrique et pluviométrique, PhD thesis, 280 pp, AgroParisTech - Irstea (Antony).

Lombardi, L., E. Toth, A. Castellarin, A. Montanari, and A. Brath (2012), Calibration of a rainfall–runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime, *Physics and Chemistry of the Earth, Parts A/B/C*, *42–44*, 77-84.

Mathevet, T., C. Michel, V. Andréassian, and C. Perrin (2006), A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, in *Large sample basin experiments for hydrological model parameterisation: Results of the Model Parameter Experiment - MOPEX*, edited by V. Andréassian, A. Hall, N. Chahinian and J. Schaake, pp. 211-219, IAHS Publ. 307.

Melsen, L. A., A. J. Teuling, S. W. van Berkum, P. J. J. F. Torfs, and R. Uijlenhoet (2014), Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification, *Water Resour. Res.*, *50*(7), 5577-5596.

Michel, C., C. Perrin, and V. Andréassian (2003), The exponential store: a correct formulation for rainfall-runoff modelling, *Hydrol. Sci. J.*, *48*(1), 109-124, DOI: 110.1623/hysj.1648.1621.1109.43484.

Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins?, *Water Resour. Res.*, *43*(5), W05434.

Montanari, A., et al. (2013), “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022, *Hydrol. Sci. J.*, *58*(6), 1256-1275.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models. Part I - A discussion of principles., *J. Hydrol.*, *10*(3), 282-290.

Oberlin, G., G. Galéa, and J. T. Toni (1973), Intérêt des jaugeages épisodiques pour l’estimation des étiages de petits bassins non équipés, *Cahiers de l’ORSTOM, série Hydrologie*, *X*, 349-368.

Oudin, L., V. Andréassian, T. Mathevet, C. Perrin, and C. Michel (2006), Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, *Water Resour. Res.*, *42*(7), W07410, doi: 07410.01029/02005WR004636.

Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity and ungaged catchments: confrontation on 913 French catchments, *Water Resour. Res.*, *44*, W03413, doi:03410.01029/02007WR006240.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne (2005), Which potential evapotranspiration input for a rainfall-runoff model? Part 2 -

Towards a simple and efficient PE model for rainfall-runoff modelling, *J. Hydrol.*, *303*(1-4), 290-306, DOI: http://dx.doi.org/210.1016/j.jhydrol.2004.1008.1026.

Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: potential for ungauged catchments, *Water Resour. Res.*, *43*(6), W06406.

Parajka, J., A. Viglione, M. Rogger, J. L. Salinas, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins - Part 1: Runoff-hydrograph studies, *Hydrology and Earth System Sciences*, *17*(5), 1783-1795.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*(1-4), 275-289, DOI: 210.1016/S0022-1694(1003)00225-00227.

Perrin, C., V. Andréassian, C. Rojas-Serna, T. Mathevet, and N. Le Moine (2008), Discrete parameterization of hydrological models: evaluating the use of parameter sets libraries over 900 catchments, *Water Resour. Res.*, *44*, W08447, doi:08410.01029/02007WR006579.

Perrin, C., L. Oudin, V. Andréassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow knowledge on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*(1), 131-151, DOI: 110.1623/hysj.1652.1621.1131.

Plasse, J., G. Drogue, D. Francois, P. Battaglia, and C. Conan (2014), Using point flow measurements for guided reconstruction of daily streamflow time series through continuous flow simulation: a regional case study on the Rhine-Meuse district, *Houille Blanche-Revue Internationale De L Eau*(1), 45-52.

Pushpalatha, R., C. Perrin, N. Le Moine, and V. Andréassian (2012), A review of efficiency criteria suitable for evaluating low-flow simulations, *J. Hydrol.*, *420-421*, 171-182.

Rojas-Serna, C. (2005), Quelle connaissance hydrométrique minimale pour définir les paramètres d'un modèle pluie-débit ? (Which is the minimum hydrometric information to estimate the parameters of a rainfall-runoff model?), PhD thesis, 319 pp, Cemagref (Antony), ENGREF (Paris), France.

Rojas-Serna, C., C. Michel, C. Perrin, and V. Andréassian (2006), Ungauged catchments: How to make the most of a few streamflow measurements?, *IAHS Publ. 307*, 230-236.

Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, *98-99*, 279-293.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241.

Seibert, J., and K. J. Beven (2009), Gauging the ungauged basin: how many discharge measurements are needed?, *Hydrology and Earth System Sciences*, *13*(6), 883-892.

Seibert, J., and J. McDonnell (2013), Gauging the Ungauged Basin: Relative Value of Soft and Hard Data, *Journal of Hydrologic Engineering*, *20*(1), A4014004.

Singh, S. K., and A. Bardossy (2012), Calibration of hydrological models on hydrologically unusual events, *Adv. Water Resour.*, *38*, 81-91.

Sivapalan, M. (2003), Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrol. Processes*, *17*(15), 3163-3170.

Tada, T., and K. J. Beven (2012), Hydrological model calibration using a short period of observations, *Hydrol. Processes*, *26*(6), 883-892.

Temnerud, J., J. Seibert, M. Jansson, and K. Bishop (2007), Spatial variation in discharge and concentrations of organic carbon in a catchment network of boreal streams in northern Sweden, *J. Hydrol.*, *342*(1-2), 72-87.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521-2535.

Vidal, J.-P., E. Martin, L. Franchisteguy, M. Baillon, and J.-M. Soubeyroux (2010), A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *International Journal of Climatology*, *30*(11), 1627-1644.

Viviroli, D., and J. Seibert (2015), Can a regionalized model parameterisation be improved with a limited number of runoff measurements?, *J. Hydrol.*, *529, Part 1*, 49-61.

Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi: 1310.1029/2001WR001118.

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

Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422.

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.

List of Tables

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.

Figure captions

Author-produced version of the article published in Water Resources Research (2016), vol. 52, n° 6 p. 4765-4784
The original publication is available at http://onlinelibrary.wiley.com/doi/10.1002/2015WR018549/epdf ©.